UNIVERSITY COLLEGE LONDON

Bayesian molecular clock dating and the divergence times of angiosperms and primates

Jose Antonio Barba-Montoya

Department of Genetics, Evolution and Environment, UCL Biosciences

July 2017

A thesis submitted to the University College London for the degree of Doctor of Philosophy

Declaration

I, Jose Antonio Barba-Montoya, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Specifically, two chapters of this thesis involve joint work which is specified below:

Chapter 3. This chapter is a modification of the work of Barba-Montoya et al. *New Phytol* (2017 submitted) and includes contributions by Dr. Mario dos Reis (M.d.R), Prof. Harald Schneider (H.S.), Prof. Philip Donoghue (P.C.J.D) and Prof. Ziheng Yang (Z.Y). J.B.-M., M.d.R., P.C.J.D. and Z.Y. conceived the project and designed the analysis. P.C.J.D. and H.S. complied the fossil dataset for the calibration points. J.B.-M. prepared the data sets and carried out the analyses. J. B.-M. and P.C.J.D. wrote the main draft of the manuscript. All authors contributed to the interpretation of results and worked on the manuscript.

Chapter 4. This chapter is a modification of the work of Barba-Montoya et al. *Mol Phylogenet Evol* (2017) and includes contributions by M.d.R, and Z.Y. M.d.R, and Z.Y. conceived the project and designed the analysis. J.B.-M. prepared the data sets and carried out the real data analysis. M.d.R carried out the theoretical 5-species analysis. All authors contributed to the interpretation of results and worked on the manuscript.

All photographs and figures by the author unless otherwise indicated.

In addition to the above, during my PhD studies I collaborated on the following paper: dos Reis M, Gunnell GF, Barba-Montoya J, Wilkins A, Yang Z. and Yoder A. Submitted 2017. Establishing a timeline of Primate divergences using genomic datasets. *Syst Biol.* See Appendix D.

Abstract

The explosive increase of molecular sequence data has produced unprecedented opportunities for addressing a number of evolutionary problems. Specially, the species divergence time estimation is fundamental because our understanding of history of life depends critically on knowledge of the ages of major clades. This thesis explores the use of molecular data (genome-scale datasets), combined with statistical summaries of the fossil record, to date the origin of angiosperms (flowering plants) and the divergence times of its major groups in an attempt to resolve the apparent conflict between the molecular dates and fossil evidence. Moreover, because fossil calibrations are the major source of information for resolving the distances between molecular sequences into estimates of absolute times and absolute rates in molecular clock dating analysis, several strategies for converting fossil calibrations into the prior on times are evaluated. Chapter one introduces the diversity and evolution of angiosperms, reviews the current literature that is based predominantly on systematics, phylogenetics, palaeobotany and plant molecular evolution. In introducing the early evolution of angiosperms this chapter highlights the questions associated with the origin of angiosperms and presents aims of the thesis. Chapter two focuses on molecular clock dating methods. It discusses different approaches for estimating divergence times, with emphasis on Bayesian molecular clock dating methods. Chapter three uses a powerful Bayesian method to analyze a molecular dataset of 83 genes from 644 taxa of vascular plants, combined with a suite of 52 fully-justified fossil calibrations to disentangle the pattern of angiosperm diversification. The results indicate that crown angiosperms originated during the Triassic to the Jurassic interval, long prior to the Cretaceous Terrestrial Revolution. This analysis demonstrates that even though many sources of uncertainty are explored, attempts to control for these factors still do not bring clock estimates and earliest confident fossil occurrences into agreement. A post-Jurassic origin of angiosperms was rejected, supporting the notion of a cryptic early history of angiosperms. The main factors affecting the estimates in this study are also discussed. Subsequently, in chapter four different strategies for summarizing fossil information to construct calibration priors were assessed employing an a priori procedure for deriving accurate calibration densities in Bayesian divergence dating.

In general, truncation has a great impact on calibrations so that the effective priors on the calibration node ages after the truncation can be very different from the user-specified calibration densities. The different strategies for generating the effective prior also had considerable impact, leading to very different marginal effective priors. Arbitrary parameters used to implement minimum-bound calibrations were found to have a strong impact upon the prior and posterior of the divergence times. The results highlight the importance of inspecting the joint time prior used by the dating program before any Bayesian dating analysis. Finally, chapter five draws together key finding from chapters three and four, and reviews how this work advances our understanding of the origin and evolution of angiosperms and on molecular clock dating using fossil calibrations. This chapter also highlights new gaps in our understanding of early evolution of angiosperms and in the implementation of fossil calibrations in Bayesian molecular clock dating, and discusses several areas for future research. Overall, this thesis highlights that more room for improvement might lie in refining our knowledge and use of fossil calibrations, the resulting improvements to molecular estimates of timescales will lead to a better understanding of angiosperm evolution. I speculate that these results will also shed light on dating discrepancies in other major clades.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Ziheng Yang for his guidance, support and patience during my PhD studies. His immense knowledge, rigour and interest in science will remain a model to me.

I would also like to thank Mario dos Reis for his insightful comments and encouragement during my studies and throughout our collaborative research. Also Prof. Phil Donoghue and Prof. Harald Schneider for sharing their knowledge and expertise so willingly. I am grateful as well to my fellow labmates Konstantinos Angelis, Daniel Dalquen, Chengmin Shi and Yuttapong Thawornwattana for many stimulating scientific discussions and useful advice, and to Prof. Max Telford for being my second supervisor.

I would like to particularly acknowledge CONACyT-México and UCL for the financial support without which this work would not have been possible. I am also indebted to the Biotechnology and Biosciences Research Council, UK (BBSRC) and the Natural Environment Research Council, UK (NERC) for providing financial support necessary to perform the research presented in chapters three and four.

To my parents Tere and Antonio and my sister Adriana, for without their almost unbelievable support and encouragement this thesis would not have been possible. Finally, I would like to thank my partner Cecilia for her constant support and patience all these years. Thank you for believing in me.

Impact statement

The timing of origination and diversification of angiosperms has been an old, and quite controversial topic in evolutionary biology. The relatively late appearance of the angiosperms in the fossil record, especially in comparison with other major groups of seed plants, suggests that the group diversified relatively recently. However, the resolution of when did angiosperms initiate the diversification that lead to their present-day species diversity has remained unclear. The oldest unambiguous angiosperm fossil extends back to the early Cretaceous, conservatively ~ 126 MY ago, but, recent efforts to date the origin of the angiosperms based molecular data have converged on much older estimates. In this study, we used a powerful Bayesian statistical method, together with the largest taxon-locus dataset ever assembled of angiosperms and a comprehensive set of fossil calibrations to address the timescale of angiosperm diversification. We have gone even a step further and have provided several interpretations of the fossil data, which gives us a broad picture of the relationship between fossil uncertainty and divergence time estimation. This timescale represents a framework for further investigating angiosperm evolution, for example, the rate molecular and morphological change, biogeographical history, diversification dynamics, ancestral character reconstruction and state-dependent diversification; as well as coevolution with other biological lineages, correlations between diversification and the physical environment, and the evolution of modern terrestrial ecosystems. Our Analysis may also shed light on dating discrepancies in other major clades.

Fossil calibrations are of particular importance for molecular clock dating since it is not possible to obtain time estimates based solely on molecular data. In a subsequent study, we examine how different calibration strategies, the birth-death process, and automatic truncation interact to determine the time prior. We show that the different calibration strategies as well as truncation have significant impacts on the time prior and the resulting posterior time estimates. Our analyses highlight the arbitrary nature of the procedure used by dating software to generate the time prior from the same fossil calibration information. We discuss the implications of our results and give recommendations for the construction of reasonable time priors. The study presented here is complementary to the existing research on molecular clock dating, which also

is an important tool in medical and clinical research. For example, it can be used to infer heterogeneous tissue age among patients with Barrett's esophagus (Curtius et al., 2016), or to study disease transmission, and virus spreading and origination (Stadler 2009). Molecular clock dating can be also used in industry, for example in brewing and wine production, which are among the oldest technologies and their products are almost indispensable in our lives. The central biological agents of beer and wine fermentation are yeasts belonging to the genus *Saccharomyces*, which can accumulate ethanol. Molecular clock dating has made it possible to elucidate when and why yeasts produce ethanol in high concentrations, and how this remarkable trait originated and developed during their evolutionary history (Piskur et al., 2006).

Table of contents

Table of figures

Table of tables

1 Introduction to angiosperm diversity and evolution

1.1 General characteristics of angiosperms

Angiosperms or flowering plants are the largest and more diverse group of extant land plants (embryophytes).They are among the most successful organisms in the history of life, representing one of the largest branches of the tree of life, as they include around 295,000 described species (Christenhusz & Byng 2016). However the estimated number of extant species may exceed 352,000 (Paton et al., 2008). Life on Earth today is extremely associated to angiosperms and there can be no doubt that angiosperms have influenced the composition and function of modern terrestrial life, from their role as primary producers, their influence on weathering and global biochemical cycles, to the creation of environments habitable by other organisms (Schneider et al., 2004; Bell et al., 2010; Magallón et al., 2015). Angiosperms are fundamental structural components in modern terrestrial ecosystems, and contribute huge amount of diversity in species richness and functional innovations (Magallón et al., 2015) (Figure 1.1). With this exceptional diversity and abundance angiosperms have enabled the development of rich and complex long-term evolutionary interactions within and among trophic levels and have promoted the diversification of other lineages (e.g., bacteria, fungi, insects and vertebrates) (Magallón & Castillo 2009). Angiosperms are also ecologically diverse; they are found in almost every terrestrial habitat, as well as in both fresh and saltwater (except in the highest mountaintops, the regions immediately surrounding the poles and the deepest oceans) (Soltis et al., 2004; Bell et al., 2010) (Figure 1.1).

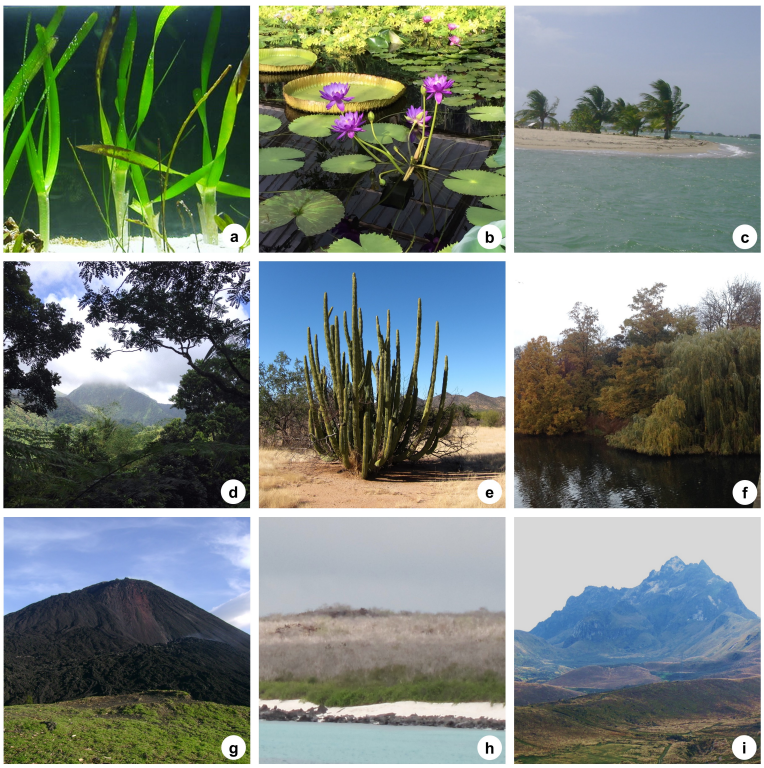


Figure 1.1: The wide range of angiosperm habitats. (a) *Thalassia testudinum,* a seagrass which grows in marine, fully saline environment; **(b)** *Nymphaea capensis,* aquatic plant which is found growing abundantly in freshwater habitats in Africa and as an introduced species in other tropical areas such as Australia and Flordia; **(c)** palm trees (Arecaceae), growing in seashore in Placencia, Belize; **(d)** a dense tropical forest in Martinique; **(e)** *Stenocereus thurberi,* a cactus growing in the Sonaran Desert, Mexico, the hottest desert in North America; **(f)** *Salix alba, Acer sp, and Quercus sp*., growing in a temperate deciduous forest in the UK; **(g)** grasses (Poaceae) growing on Pacaya, an active volcano in Guatemala; **(h)** *Bursera graveolens, Cryptocarpus sp*., and *Sesuvium sp*., growing in an isolated island in Galapagos, Ecuador; **(i)** grasses (Poaceae) growing on slopes of the high Rucu summit of "The Pichinchas" in Ecuador.

Angiosperms display a great variety of life forms including floating and rooted submerged aquatics, epiphytic, lithophytic and terrestrial plants that differ greatly in longevity, size and growth form. They can be gigantic trees, shrubs, small herbs, palm-like plants, bulbs, vines, lianas, cushion plants, rosettes, canes or parasitic plants (Du Rietz 1931; Adamson 1939) (Figure 1.2). Angiosperms not only comprise an extraordinary vast morphological and phylogenetic diversity, there is also a considerable amount of diversity in biochemistry, physiology, reproductive morphology, and genome size and organisation that is unparalleled in other members of the Plantae Kingdom (Soltis & Soltis 2004). Furthermore, angiosperms are essential for the human being; they have high economic and cultural values; not only the majority of the world's crops are angiosperms but they are also sources of

important resources such as medicine, timber, natural clothing fibres, paper and decorative and landscaping plants (Soltis et al., 2004; Reyes-Garcia et al., 2005).

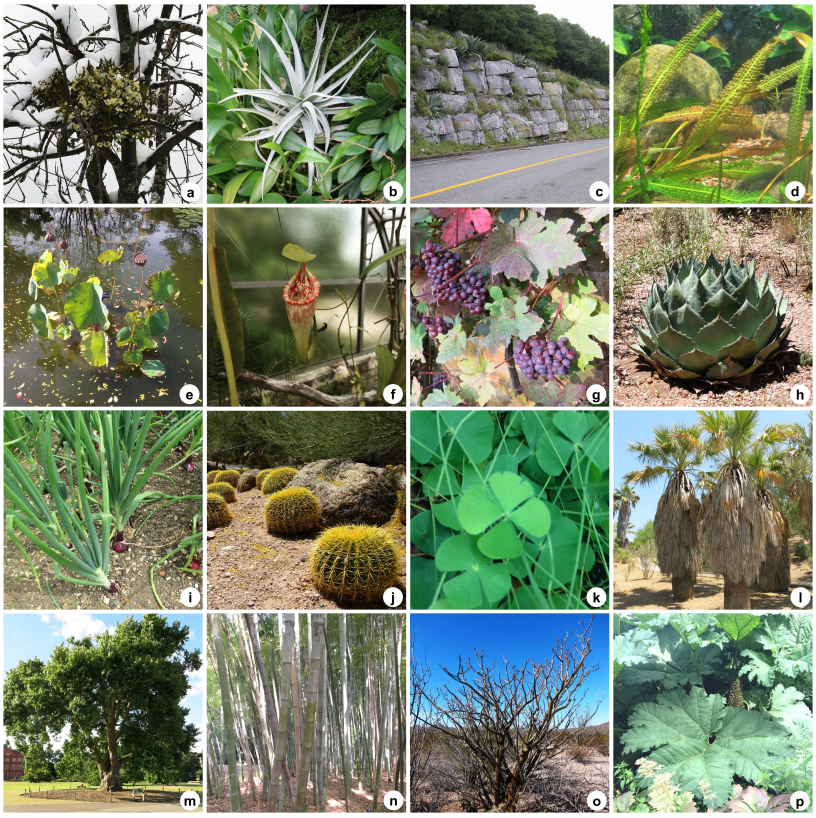


Figure 1.2: Angiosperm life-forms. (a) parasite, *Viscum album*; **(b)** epiphyte, *Tilandsia sp.*; **(c)** lithophyte, agaves and grasses; **(d)** submerged aquatic, *Aponogeton boivianus*; **(e)** floating aquatic, *Nelumbo sp.*; **(f)** liana, *Nephenthes sp.*; **(g)** vine, *Vitis sp.*; **(h)** rosette, *Agave parrasana*; **(i)** bulb, *Allium cepa*; **(j)** barrel-form, *Echinocactus grusonii*; **(k)** small herb, *Trifolium sp.*; **(l)** palm-like, Arecaceae fam.; **(m)** tree, *Platanus orientalis*; **(n)** cane, bamboo (Poales); **(o)** shrub, *Fouquieria sp.*; **(p)** large leaves, *Gunnera sp*. Figure (a) adapted from Wikipedia.

Notwithstanding their high diversity, angiosperms are certainly unified by a set of synapomorphies or shared derived features including 1) ovules that are enclosed within a carpel, a structure made up of an ovary, which encloses the ovules and the stigma, a structure where pollen germination takes place; 2) formation of an endosperm leaded by double fertilization (this highly nutritive tissue is formed within the seeds and feeds the developing plant embryo, the cotyledons and the seedlings when they first appear); 3) stamens with two pairs of pollen sacs; 4) features of gametophyte structure and development (this is significantly reduced in size compared to those of gymnosperm seed plants, is composed by three cells in males and by seven cells and eight nuclei in females); and 5) phloem tissue composed of

sieve tubes and companion cells. All available evidence, including molecular data, strongly support the monophyly of extant angiosperms (Doyle & Donoghue 1986; Judd et al., 2002; Soltis et al., 2004; Magallón 2009).

1.2 Systematics of angiosperms

As mentioned in the previous section, angiosperms constitute a monophyletic group very well supported by molecular data and by a large number of unique traits. They are classified into 64 orders, 416 families and approximately 13,164 genera (APG_IV 2016; Christenhusz & Byng 2016). Angiosperms are seed-bearing vascular plants (Figure 1.3). Their reproductive structures are flowers in which the ovule is fertilised and develops into a seed enclosed in a hollow ovary. Fruits are derived from the maturing floral organs and contain the seeds. A solid understanding of the phylogenetic affinity of the majority of living angiosperms and of relationships among clades has been achieved by the botanical community (Cantino et al., 2007; Soltis et al., 2011; Ruhfel et al., 2014; APG_IV 2016).

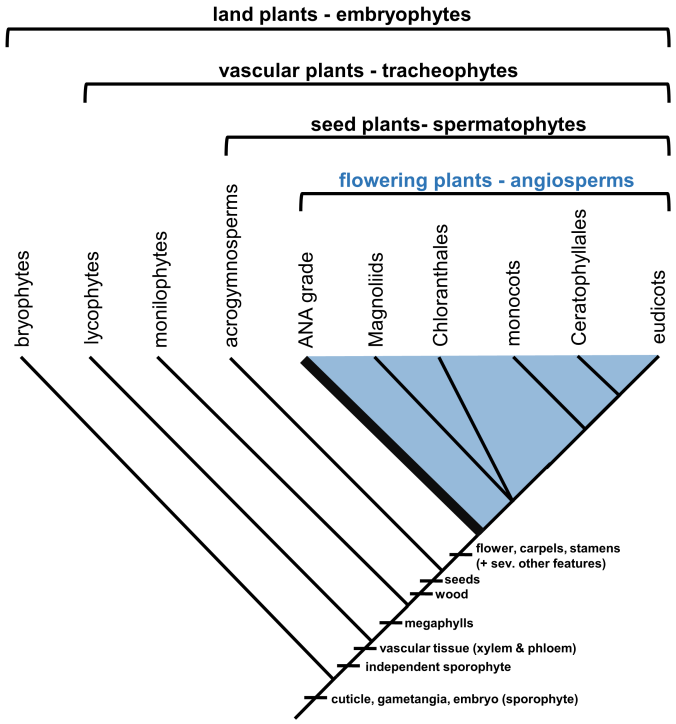


Figure 1.3: Embryophyte phylogeny. A simplified evolutionary tree of the land plants representing the major groups and showing synapomorphyc characters of each group (Simpson 2006; Cantino et al., 2007; Soltis et al., 2011; Ruhfel et al., 2014; APG_IV 2016). Angiosperms are highlighted in blue with the major groups (ANA grade, Magnoliids Chloranthales, monocots, Ceratophyllales and eudicots) represented within the clade.

Angiosperm phylogeny is relatively well known compared to many other groups of organisms, although a few families and genera remain to be phylogenetically placed. The living angiosperms are distributed in eight major lineages (APG_IV 2016) (Figures 1.3 and 1.4). Amborellales (1 sp.), Nymphaeales (88 spp.) and Austrobileyales (94 spp.) are the earliest branches and form the ANA grade, which encompass only ~ 0.05% of their standing species richness. The vast majority of the living angiosperms belong to a lineage referred to as core angiosperms (Mesangiospermae), which includes Chloranthales (77 spp.), magnoliids (10,842 spp.), monocots (74,273 spp.), Ceratophyllales (5 spp.) and eudicots (210,008 spp.). The deep-level relationships in the angiosperms are being resolved with strong support. Amborellaceae, Nymphaeales, and Austrobaileyales are successive sisters to the remaining angiosperms. Chloranthales is placed on a polytomy with magnoliids and sister to monocots/Ceratophyllales/eudicos. Eudicots are classified into 4 subgroups: the basal eudicots grade (subtending the Rannunculales, Proteales, Trochodendrales, Buxales and Gunnerales), the superrosids, the superasterids and the Dilleniales (recently placed in a polytomy with superrosids and superasterids) (APG_IV 2016; Christenhusz & Byng 2016).

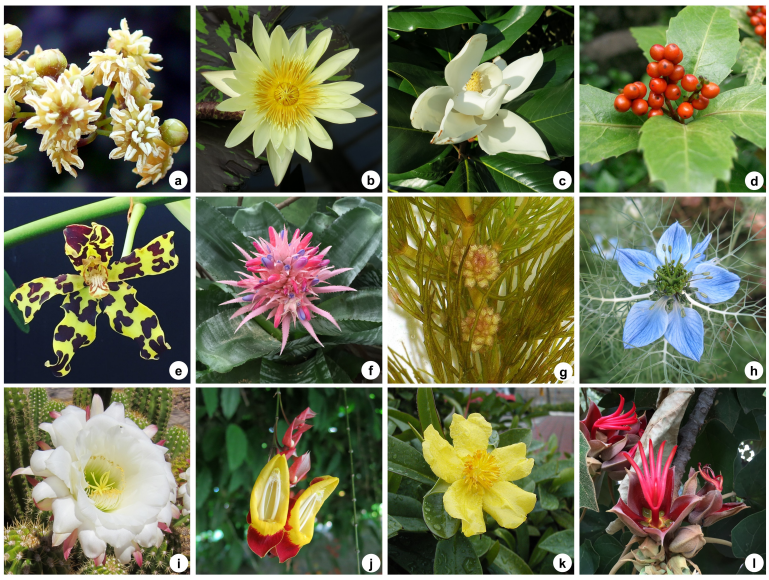


Figure 1.4: Angiosperm diversity. (a) *Amborella trochopoda* (ANA grade); **(b)** *Nymphaea sp.* (ANA grade); **(c)** *Magnolia grandifolia* (magnoliids); **(d)** *Sarcandra glabra* (Cloranthales); **(e)** *Grammatophyllum scriptum* (monocots); **(f)** *Aechmea Fasciata* (monocots); **(g)** *Ceratophyllum demersum* (Ceratophyllales); **(h)** *Nigella damascena* (Ranunculales); **(i)** *Echinopsis sp* (eudicots, superasterids); **(j)** *Thunbergia mysorensis* (eudicots, asterids); **(k)** *Hibbertia scandens* (Dilleniales); **(l)** *Chiranthodendron pentadactylon* (rosids). Figures (a), (d), (g) and (k) adapted from Wikipedia.

1.3 Origin and timing of angiosperm diversification

1.3.1 Hypothesis of seed plant relationships

One essential aspect of phylogeny for understanding angiosperm origin is their relation to other seed plants. For more than 150 years, attempts to understand the origin and diversification of angiosperms were hampered by uncertain relationships among the great diversity of extant angiosperms, as well as the seemingly insurmountable morphological gap between angiosperms and other seed plants (gymnosperms) (Friis et al., 2011). It should be clear that angiosperm crown group is defined as the least inclusive monophyletic group that includes the most recent common ancestor of extant angiosperms and all its derivatives. Angiosperm stem group is the most inclusive monophyletic group containing the extant members of the clade but no other extant taxa. It also includes the extinct lineages that diverged from the lineage leading to the angiosperm crown group (angiosperm stem relatives). These must all be more closely related to the angiosperm crown group than to any other extant group (Figure 1.5) (Doyle & Donoghue 1993; Magallón & Sanderson 2001).

Over the past few decades in various cladistic analyses, the extant sister group of the angiosperm crown group has been identified to be either Gnetophyta (the anthophyte hypothesis) (Crane 1985; Doyle & Donoghue 1986; Doyle et al., 1994; Rothwell & Serbet 1994; Hilton & Bateman 2006), a clade comprising *Gnetum* and *Welwitschia* (Nixon et al., 1994), a clade comprising conifers and Gnetophyta (Hill & Crane 1982; Soltis et al., 2002), a clade comprising conifers, cycads and *Ginkgo* (Magallón & Sanderson 2002; Rydin et al., 2002; Rai et al., 2003), or Cycadophyta (Doyle 2006). However, the consensus emerging from phylogenetic analyses of molecular data (Bowe et al., 2000; Magallón & Sanderson 2002; Soltis et al., 2002; Cantino et al., 2007; Soltis et al., 2011; Magallón et al., 2013; Ruhfel et al., 2014; Magallón et al., 2015; Barba-Montoya et al., Submitted 2017) clearly shows that Acrogymnospermae (name for the clade that includes all extant gymnosperms) is a monophyletic group and is sister of all angiosperms.

1.3.2 Origin and early history of angiosperms

Outgroup relationships may provide valuable evidence for understanding the evolution of several distinctly angiosperm organs such as flowers, leaves, stamen, ovule, carpel and the vegetative structure. Also, resolution of the relationships among angiosperms themselves, together with palaeobotanical information, allows some conclusions to be drawn about the origin of some distinctive angiosperm features (Friis et al., 2011). In the past two decades, phylogenetic analyses of molecular data have transformed our thinking about the origin and evolution of angiosperms (Doyle 2012). As insights from work on living and fossil angiosperms have continued to accumulate, some hypotheses or scenarios have been developed regarding the form of the ancestral angiosperm. Although some morphological cladistic results have been refuted by molecular analyses, others that were debatable at first have been confirmed. The anthophyte hypothesis, the dominant concept 30 years ago, has been discarded; Gnetales, once thought to be the closest living relatives to angiosperms, are actually related to other extant gymnosperms (acrogymnosperms), most likely associated to conifers (Frohlich & Chase 2007). Moreover, the cladistic approach resolved evolutionary questions by formulating more explicit hypotheses on phylogeny and its relation to character evolution, which is no less important in the present molecular era. Results from molecular data analyses concern only crown groups (Doyle 2012).

Finding fossils related to crown angiosperms or their stem relatives can further the understanding of homologies among these groups; for instance, morphological analyses of living and fossil seed plants that assume molecular relationships recognize glossopterids, Bennetitales and *Caytonia* as angiosperm relatives (Figure 1.5) (Doyle 2008, 2012). As previously indicated (Section 2.3.1), recent molecular data analyses (Soltis et al., 2011; Ruhfel et al., 2014; APG_IV 2016) have consistently pointed to the ANA taxa as basal angiosperms and successive sister groups to the larger clades of magnoliids, eudicots and monocots. The taxa that compose the ANA grade are each individually very specialised. For example, *Amborella* (a single species of shrub endemic to cloud forest of New Caledonia) lives naturally in wet, forest understorey habitats and is dioecious with vestigial organs of the opposite sex, whereas Nymphales (a cosmopolitan clade) with perfect flowers are all adapted to aquatic habitats and Austrobraleyales are woody aromatic plants growing as trees,

shrubs and lianas (Soltis et al., 2004; Frohlich & Chase 2007). Based on parsimony analysis of character evolution Doyle (2012) inferred that the most recent common ancestor of angiosperms had (1) vesselless wood, (2) pinnately veined simple leaves, (3) monosculate pollen with columellar exine structure, (4) more than two whorls or series of undifferential perianth parts (tepals), (5) numerous stamens, and (6) more than one ascidiate carpel containing a single pendent bitegmic ovule (Doyle 2005; Doyle 2007; Doyle 2009; Endress & Doyle 2009).

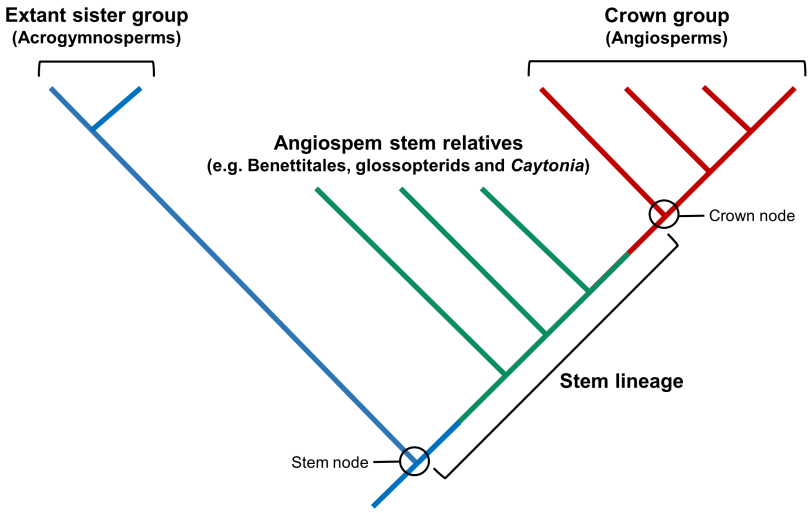


Figure 1.5: Phylogeny of angiosperms and related seed plants. Crown group angiosperms (red) possess all the apomorphies of the extant taxa. Lineages in the angiosperm stem group (green) have acquired some, but not all, off the apomorphies of the angiosperm crown group. Acrogymnosperms represent the extant sister group (blue) of all the angiosperms. Adapted from Doyle (2012).

In relation to the age of the angiosperms, it is important to distinguish two dates. The age of the angiosperm crown group corresponds to the time of the first phylogenetic split within the crown group. The age of the angiosperm stem group, which may be significantly older, corresponds to the time of divergence between the angiosperm crown group and its extant sister group (acrogymnosperms) (Magallón & Sanderson 2001). The angiosperm crown group could have originated anywhere between the divergence of the stem lineage and the first crown group fossils in the Lower Cretaceous (Brenner & Bickoff 1992; Brenner 1996; Taylor & Hickey 1996; Friis et al., 2000a; Clarke et al., 2011; Doyle 2012). Molecular clock dating provides a way of estimating the absolute time between the stem and crown nodes. Meanwhile, identification of fossils with morphologies that convincingly place them close to the angiosperm stem or crown groups could revolutionize our understanding of the origin

of the angiosperms (Frohlich & Chase 2007; Doyle 2012). Section 1.3.3 provides a brief review of the age of the angiosperms.

Moreover, new insights into the early flower and the early angiosperm diversification have been recently provided based on genome structure (Adams 2013; Amborella_Genome_Project 2013), gene function (Zahn et al., 2005; Frohlich & Chase 2007; Soltis et al., 2008) and whole-genome duplication (WGD) (Jiao et al., 2011; Li et al., 2015; Wendel 2015).

The genome sequence of *Amborella trichopoda,* the sister taxon to all other living angiosperms, provides insights into the molecular evolution of angiosperms and has major implications for reconstructing characteristics of an ancestral angiosperm genome. For instance, it facilitates inference of the gene content and structure of the earliest angiosperms (Adams 2013). The ancestral angiosperm gene set contained more than 14,000 protein coding genes. In relation to nonangiosperm seed plants (spermatophytes), 1179 gene lineages first appeared in association with the origin of angiosperms. These new gene lineages may have led to gene functions specific to angiosperms and critical to their early diversification (Adams 2013; Amborella_Genome_Project 2013).

In addition, novel functions were achieved by pre-existing genes. For example, the genes involved in flower development that have homologs in other spermatophytes (Adams 2013). MADS-box transcription factors are essential for flower origins because they specify the major floral organs. According to the ABCE model for floral-organ pattering (Coen & Meyerowitz 1991; Zahn et al., 2005), the floral-organ identity is controlled by four gene functions, A, B, C, and E, that operate in combination to produce floral organs. Sepals are specified by A-function alone, petals by A-function in combination with B-function, stamens by B-function in combination with C-function, carpels by C-function alone. E-function are required for all four organ types (Soltis et al., 2008). Ovules are specified by D-function. Phylogenetic analysis demonstrate that each major MADS subgroup extends back to the base of crown angiosperms (Zahn et al., 2005; Frohlich & Chase 2007).

Several studies provide evidence that an ancient WGD predated angiosperm diversification (Jiao et al., 2011; Amborella_Genome_Project 2013; Li et al., 2015).

WGD followed by gene loss and diploidization has long been recognised as a common mode of speciation and evolution in angiosperms, so that their success has been ascribed partly to innovations related to gene or whole genome duplications (Jiao et al., 2011). The ancestral angiosperm was a polyploid with an enormous set of novel and ancient genes that survived to play crucial roles in angiosperm biology (Amborella_Genome_Project 2013).

Detecting genomic changes that co-occurred with the origin of the angiosperms is fundamental to elucidating the molecular basis of biological innovations that contributed to their rise to ecological predominance (Amborella_Genome_Project 2013). Further research on gene function and genetic mechanisms that control reproductive development in spermatophytes may allow us to define how the first angiosperm was and why they evolved this way (Frohlich & Chase 2007).

1.3.3 The age of the angiosperms

1.3.3.1 The fossil record

Only a small proportion of the global vegetation will ever leave evidence of its existence in the fossil record; moreover, the processes responsible for the formation of plant fossils are complex (Friis et al., 2011). However, many of the biases of representation and recognition inherent in the plant fossil record have been clarified, yet there are still difficulties in identifying fossils' affinities that have not been completely resolved and that have significant implications for determination of timing in angiosperm evolution with molecular clock models (Crepet 2008; Friis et al., 2011). Moreover, it is useful to distinguish between the time of divergence of the angiosperm clade from its living sister taxon (the stem group age) from the time of the oldest divergence that gave rise to a living angiosperm species (the crown group age). The principal sources of information on the timing and pattern of early angiosperm diversification are studies of detached leaves, dispersed pollen, well-preserved flowers and floral organs (Crepet 2008; Friis et al., 2011; Doyle 2012).

The sequence appearance of major angiosperm pollen types suggests that the first angiosperm pollen was monosulcate (characterised by a single furrow for pollen tube germination), followed by tricolpate pollen (characterised by three longitudinal

furrows). Nevertheless, monosulcate pollen grains from the Valanginian (Brenner & Bickoff 1992; Brenner 1996) and Hauuterivian to Barremian (Huges & McDougall 1987; Huges 1994) may not be restricted to the angiosperm crown group. Therefore, tricolpate pollen is the oldest unequivocal fossil record of crown-angiosperms at ~126 million years ago (Ma) (Friis et al., 2000b). The earliest unambiguous evidence of angiosperms, based on Fischer's rule tricolpate pollen, can be constrained minimally to the Barremian, but this actually evidences the establishment of the eudicot lineage, which is remote from the angiosperm crown ancestor (Clarke et al., 2011).

Angiosperms rose to ecological predominance in the Cretaceous Terrestrial Revolution (KTR), from 125-80 Ma (Lloyd et al., 2008), when their apparently explosive radiation is believed to have fostered the diversification of lineages that are main contributors to extant terrestrial environments such as pollinating insects, leaf-eating flies, butterflies, moths, squamates (lizards and snakes), crocodilians, basal groups of placental mammals and modern birds, establishing the modern terrestrial biodiversity (Dilcher 2000; Benton 2010; Meredith et al., 2011; Augusto et al., 2014; Cascales-Minana et al., 2016). Nevertheless, these hypotheses of co-diversification are mostly based on the perceived coincidence in the radiation of angiosperms and the renewal of trophic networks in terrestrial ecosystems. This is evidenced, not least, by the fossil record of tricolpate pollen in the Barremian, slightly younger Aptian floral assemblages, followed by an explosive increase in diversity in the middle and late Cretaceous (Doyle 2008; Clarke et al., 2011; Magallón et al., 2015; Herendeen et al., 2017). Some interpret this evidence literally to reflect an explosive radiation from a Cretaceous crown ancestor, with the earliest macrofossil record of an unambiguous angiosperm, (Friis et al., 2000b; Sun et al., 2002) dating back only to the mid-Early Cretaceous (Hickey 1997; Dilcher 2000; Benton 2010; Friis et al., 2010; Meredith et al., 2011; Doyle 2012; Gomez et al., 2015; Cascales-Minana et al., 2016; Herendeen et al., 2017).

Monosulcate pollen, like that produced by early-branching lineages of extant angiosperms, is known at least as far back as the Valanginian (Brenner 1996), and pollen exhibiting subsets of definitive crown-angiosperm characters are known as far back as the Middle Triassic (Cornet 1989; Doyle & Hotton 1991; Taylor & Taylor 2009; Hochuli & Feist-Burkhardt 2013). These, however, are difficult to discriminate from

pollen produced by stem-angiosperms or gymnosperms (Hochuli & Feist-Burkhardt 2013) and, hence, they have not been used to constrain divergence time analyses. There are also claims of pre-Cretaceous crown-angiosperms based on macrofossil evidence. While the age of the angiosperm macrofossil genus *Archaefructus* (Sun et al., 2002; Friis et al., 2003) has been revised from Jurassic to Cretaceous (Chang et al., 2009). Other putative pre-Cretaceous angiosperm fossils are more securely dated but their interpretation needed further attention (Crane et al., 1995; Taylor & Taylor 2009; Friis et al., 2011; Doyle 2012; Liu & Wang 2016a; Liu & Wang 2016b; Herendeen et al., 2017). New Jurassic records still needed to be further scrutinised (Liu & Wang 2016a; Liu & Wang 2016b). At best, these hypotheses of character evolution should be reviewed within a probabilistic framework that can better accommodate the uncertainty associated with such inference. Nevertheless, it may be more appropriate to reconsider the phylogenetic position of critical fossil taxa using likelihood models of character evolution to accommodate phylogenetic affinity (O'Reilly et al., 2016), since discriminating between a stem- and crown-angiosperm affinity of all pre-Cretaceous claims may be the only way in which molecular estimates for the origin of flowering plants will be more precise.

1.3.3.2 Molecular age estimates for the angiosperms

Estimates of the age of the angiosperms and the timing of important divergences have been the focus of numerous molecular clock studies (e.g. Ramshaw et al., 1972; Martin et al., 1993; Sanderson & Doyle 2001; Schneider et al., 2004; Bell et al., 2005; Bell et al., 2010; Magallón 2010; Smith et al., 2010; Clarke et al., 2011; Magallón et al., 2013; Zanne et al., 2014; Beaulieu et al., 2015; Magallón et al., 2015; Foster et al., 2016). Apart from a wide range of methodological differences, interpretation of paleontological data and implementation of fossil constraints can account for many differences between the age estimates obtained in these studies. Such studies are based on different types of molecular data, are conducted under different molecular clock methods, include differential taxonomic density and sampling, and implement different calibrations and time constraints.

Molecular clock estimates of crown angiosperms range between 300 Ma or older (Ramshaw et al., 1972; Martin et al., 1993; Magallón 2010) and 68 Ma (Sanderson & Doyle 2001), with most estimates lying between c. 140 and 243 Ma (Zanne et al.,

2014; Beaulieu et al., 2015; Magallón et al., 2015; Foster et al., 2016). These recent studies have invariably concluded that crown angiosperms diverged as much as 115 million years (Myr) earlier than the oldest unequivocal fossil record of crown-angiosperms at ~126 Ma (Zanne et al., 2014; Beaulieu et al., 2015; Foster et al., 2016), unless they have been forced to fit within the age of the oldest fossil angiosperms (Magallón et al., 2015). Estimated ages for major angiosperm clades using molecular data are also generally older than inferences from the fossil record, but these discrepancies are generally small.

1.4 Aims of this thesis

1.4.1 Estimating the timescale of angiosperm origin and diversification — chapter three

The timing of origination and early evolution of angiosperms has been an old and quite controversial topic in evolutionary biology, from Darwin himself who considered it "an abominable mystery" (Davies et al., 2004; Crepet & Niklas 2009; Friedman 2009). Writing to his friend Joseph Hooker in 1879, Darwin famously referred to the origin of angiosperms:

> “The rapid development, as far as we can judge, of all the higher plants within recent geological time is an abominable mystery”

Nowadays, Darwin’s “abominable mystery” refers to a problem about timing of origination and early evolution of angiosperms (Crepet & Niklas 2009; Friedman 2009). A particular point of interest has been working out the pattern of angiosperm diversification during the Cretaceous and its relationship with the KTR (Dilcher 2000; Schneider et al., 2004; Benton 2010; Meredith et al., 2011; Augusto et al., 2014; Feldberg et al., 2014; Laenen et al., 2014; Cascales-Minana et al., 2016). The aim of this thesis is to disentangle the pattern of flowering plant diversification using state-of-the-art Bayesian statistical methods together with the largest molecular dataset ever assembled of angiosperms and a comprehensive, well justified, set of fossil calibrations. Furthermore, this chapter provides several interpretations of the fossil data, which gives us a broad picture of the relationship between fossil uncertainty and divergence time estimation. This study controls some methodological variables from

previous studies (e.g. low taxon sampling, insufficient outgroup sampling and very limited sequence data), while also attempting to control for other sources of error including data partitioning, parameter choice in priors for rates and times, relaxed molecular clocks, and the effect of outgroup sampling. Applying the methods to our data allow us to assess the hypothesis that the crown angiosperms originated in the Cretaceous and, as such, the extreme hypothesis of a Cretaceous Terrestrial Revolution predicated on an explosive diversification of flowering plant fully within the Cretaceous.

Figure 1.6: Charles Darwin's greenhouse. In this greenhouse at Down House Darwin conducted many botanical experiments (e.g. on plant motion and phototropism; on orchid pollination and fertilization; and on heterostyly in *Primula*).

1.4.2 Evaluation of different strategies for converting fossil calibrations into the prior of times in Bayesian molecular clock dating — chapter four

Fossil calibrations are the foremost source of information for resolving the distances between molecular sequences into estimates of absolute times and absolute rates in molecular clock dating analysis (Magallón 2004; Benton et al., 2009; Ho & Duchene 2014; dos Reis et al., 2016). The quality of calibrations is thus expected to have a major impact on divergence time estimates even if a huge amount of molecular data is available (Yang & Rannala 2006; Rannala & Yang 2007; Magallón et al., 2013). In Bayesian molecular clock dating, fossil calibration information is incorporated in the analysis through the prior on divergence times (the time prior). By evaluating several strategies for converting fossils into the prior of times, Chapter 4 aims to examine how

different calibration strategies, the birth-death process, and automatic truncation (to enforce the constraint that ancestral nodes are older than descendent nodes) interact to determine the time prior. Furthermore, the implications of the obtained results are discussed with the aim of providing recommendations for the construction of reasonable time priors.

2 Molecular clock and divergence time estimation

2.1 The molecular evolutionary clock

The molecular clock hypothesis asserts that the rate of nucleotide or amino acid sequence evolution is approximately constant over time or among evolutionary lineages. Five decades ago, it was noticed that the amino acid differences between aligned haemoglobin (Zuckerkandl & Pauling 1962, 1965), cytochrome *c* (Margoliash 1963) and fibrinopetides (Doolittle & Blomback 1964) sequences from different species were roughly proportional to the times of divergence between the species (according to paleontological data). These observations led Emile Zuckerkandl and Linus Pauling to propose the hypothesis of a *molecular evolutionary clock* in 1965. From the beginning, the molecular clock was perceived as a stochastic clock in which mutations accumulate at random intervals, although at approximately the same rate in different species, thus keeping time as a clock does (dos Reis et al., 2016). An analogy of this process would be the way in which the random decay of isotopes can be used to build an atomic clock. Moreover, as different isotopes have a specific radioactive decay rate, different proteins can have different evolutionary rates; that is, their molecular clock tick at different rates (Rannala & Yang 2013).

In the 1970s many statistical analyses were conducted to test the molecular clock hypothesis on protein sequence data (e.g. Dickerson 1971; Ohta & Kimura 1971). The general insight was that a rough linear relationship exists between the estimated number of amino acid substitutions and the divergence time. However, despite these findings, the reliability of clock and its implications for the mechanism of molecular evolution were a focus of much controversy. At that time, the synthetic theory of evolution was at its peak. It was thought that the evolutionary rate was controlled by environmental changes and natural selection. Species living in different environments, with different life stories, body sizes and generation times, must have been under different selection regimes (Graur & Li 2000; Nei & Kumar 2000; Yang 2014). Moreover, the mechanisms underlying the constant rate of amino acid substitution was unclear.

Zukekandl and Pauling justified the molecular clock by suggesting that the amino acid changes that accumulate among species are mainly those with little or no effect on the structure and function of the protein, thereby reflecting the background mutational process at DNA level (dos Reis et al., 2016). This hypothesis was developed in the neutral theory of molecular evolution (Kimura 1968; King & Jukes 1969), which holds that the majority of molecular changes in evolution are due to the random fixation of neutral or nearly neutral mutations, rather than due to fixation of advantageous mutations driven by natural selection. If molecular evolution is dominated by neutral mutations, which have little influence on fitness, then an approximately constant rate of evolution is expected, thus the molecular clock can be explained. Proteins with different functional constraints may have different proportions of amino acids experiencing neutral mutations; therefore they have different rates of neutral mutation and their clocks tick at different rates (Rannala & Yang 2013; dos Reis et al., 2016).

Further controversies exist regarding whether the neutral theory predicts rate constancy over calendar time or over generations, or whether the clock applies only to silent or DNA changes, or to protein evolution (Rannala & Yang 2013). The main factors that account for between-species rate differences are generation time, with shorter generations associated with higher substitution rate, and a DNA repair mechanism with a less reliable repair mechanism associated with higher substitution rate. Furthermore, some studies support a negative correlation between substitution rates and body size; in general species with small body size tend to have shorter generation time and higher metabolic rates (Yang 2014). Even though the rate constancy assumption has always been contentious, it has been widely used in divergence time estimation and phylogenetic reconstruction. Hence, the question of validity of molecular clock is a fundamental issue in molecular evolution (Graur & Li 2000).

2.2 Molecular clock dating

The molecular clock hypothesis provides a simple but powerful approach of measuring the timescale of evolutionary divergences. A direct implication of the clock is that the expected distance between sequences grows linearly with time of divergence. Paleontological information about the ages of at least one divergence event on a phylogeny, based on the fossil record or certain geological events (e.g.

continental breakup, age of an island), can be used to translate the distances between sequences or the branch lengths on the tree into absolute geological times. This general technique of estimating divergence times is known as molecular clock dating (Rannala & Yang 2013). The first statistical methodology of molecular clock dating that was based on distance and maximum likelihood (ML) methods assumed a constant rate of evolution over time and across lineages (strict-clock), and used fossil-age calibrations as point values, although the fossil record can never provide a precise age estimate for a specific lineage (Doolittle & Blomback 1964; Sarich & Wilson 1967a; Dickerson 1971; Ohta & Kimura 1971). Subsequent tests demonstrated that the molecular clock is often violated (Langley & Fitch 1974; Felsenstein 1981); that is, the molecular evolutionary rate is not constant, except in comparisons of closely related species with similar life history. Different factors can influence the varying rates of molecular evolution across species such as mutation rate, generation time, population size and basal metabolic rate and selection; nevertheless, the exact mechanisms of rate variation and the relative importance of these factors are still a matter of debate (Ho 2014; dos Reis et al., 2016).

Nowadays it is generally accepted that the molecular clock is not universal and does not hold to distantly related species, but might be a good approximation for closely related species, or in the analysis of population data (Near et al., 2005; dos Reis et al., 2016). When the clock is violated, methods for dealing with rate variation include: (1) pruning lineages exhibiting deviation from uniform molecular evolutionary rates (Takezaki et al., 1995), (2) using multiple models of molecular evolution with differing molecular evolutionary rates on different branches of phylogenies (local-clock models) (Rambaut & Bromham 1998; Yoder & Yang 2000), (3) using nonparametric and semiparametric models of molecular rate evolution (Sanderson 1997, 2002, 2003), and (4) using Bayesian statistical models that take into account uncertainty in the fossil record and variation in molecular evolutionary rate (relaxed-clock). These Bayesian phylogenetic models include: relaxed clock models to accommodate the violation of the clock (Thorne et al., 1998; Drummond et al., 2006; Rannala & Yang 2007); the integration of morphological characters from fossil and extant species in a combined analysis with molecular data (Pyron 2011; Ronquist et al., 2012a), as well as the integration of fossil preservation and discovery to construct prior probability distributions on node ages to be used as calibrations in molecular clock dating (Wilkinson et al., 2011). Furthermore, a special case of Bayesian molecular dating

applies to viral genes and ancient DNA — in both cases it is possible to sample sequences over an evolutionary timescale — the dates of the intermediate samples can be used to calibrate the clock and to estimate divergence times, applying similar techniques to those for dating species divergences using fossil calibrations (Yang 2014).

The outcomes of molecular clock have often been controversial, generally because the molecular dates disagree with the fossil record. Part of the discrepancy between molecular and fossil data is consequence of an incomplete fossil record (Magallón et al., 2013). Moreover, fossils provide information regarding the date by which a recently diverging lineage had developed distinctive morphological traits (i.e., a synapomorphy). There may be a lapse between the time of a lineage's origin and the age of the first fossil with the derived traits of the descendants. In contrast, molecular dating infers node ages (estimates of time of divergence events among ancestral lineages) in a phylogenetic tree. Hence, fossil-based dates are generally younger than those derived from molecular data. Besides, other sources of discrepancy can be deficiencies and inaccuracies in molecular time estimation. Despite controversies, the interplay between molecules and fossils has been a driving force in this research area, since it has encouraged much discussion about the nature of the fossil record and the molecular clock and inspired the development of more sophisticated methods (Rannala and Yang, 2013). The aim of this chapter is to describe statistical methods for testing the clock hypothesis, and likelihood and Bayesian methods of divergence time estimation under global clock and local-clock. Such methods use fossils to calibrate the clock.

2.3 Tests of molecular clock

As mentioned above, the assumption of rate constancy across lineages of a phylogeny, which would produce an approximately steady rate of accumulation of nucleotide or amino acid changes through time, has been questioned (Battistuzzi et al., 2011). Some statistical tests have been developed to examine whether rates of molecular evolution vary significantly among evolutionary lineages. Two main types of tests exist: those based on comparisons of genetic distances and those based on likelihood ratios.

The simplest type — *the relative-rate test* — examines whether two species (*a* and *b*) evolve at the same rate relative to an outgroup (species *c*). If the clock hypothesis is true, the distances from ancestral node *y* to species *a* and *b* should be equal: $d_{ay} = d_{by}$ or $b_1 = b_2$ (Figure 2.1). Similarly one can formulate the clock hypothesis relative to the outgroup as $d_{ac} = d_{bc}$ (Sarich & Wilson 1967b, 1973). A simple way to test whether an observed *d* value is significantly different from 0 is to compare it with its standard error: first, $b_1 = (d_{ab} + d_{ac} - d_{bc})/2$ and $b_2 = (d_{ab} + d_{bc} - d_{ac})/2$ are calculated as the number of changes along branches y_a and y_b, then, the clock is tested by comparing $X^2 = (b_1 - b_2)^2/(b_1 + b_2)$ against X_1^2 (Fitch 1976). Also, the binomial distribution bin($b_1 + b_2$, $\frac{1}{2}$) can be used to test whether the observed proportion $b_1/(b_1 + b_2)$ deviates significantly from $\frac{1}{2}$, with $\left(\frac{b_1}{b_1+b_2} - \frac{1}{2}\right) \Big/ \sqrt{\frac{1}{2} \times \frac{1}{2} \Big/ (b_1 + b_2)} = (b_1 - b_2)/\sqrt{b_1 + b_2}$ compared against the standard normal distribution (Yang 2014). This method fails to correct for multiple hits, and moreover, the X_1^2 and the binomial approximations may be unreliable if $b_1 + b_2$ is small. However, the sequence distances and their variances can be estimated under any model of nucleotide or amino acid substitution, and the estimated *d* and its standard error can be used to construct a test based on the normal distribution (Wu & Li 1985; Battistuzzi et al., 2011; Rannala & Yang 2013). Furthermore, the relative-rate test can also be conducted when the pattern of substitution rates is unknown and/or the substitution rate varies among different sites, or even when the outgroup is unknown. Nevertheless, this test is based on the chi-square test, like Fitch's test above (Tajima 1993).

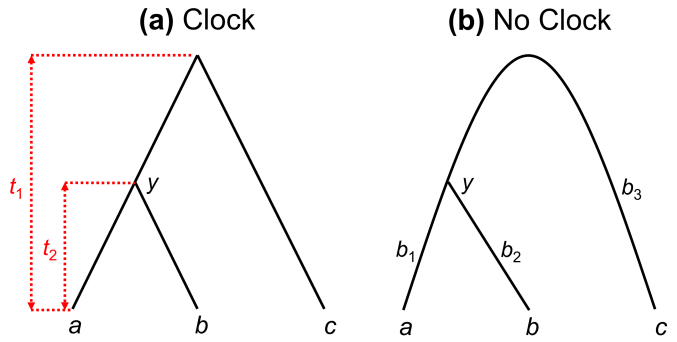


Figure 2.1: The relative-rate test. (a) Under the clock (rooted tree), the parameters are ages of two ancestral nodes t_1 and t_2, measured by the expected number of substitutions per site. **(b)** Without the clock (unrooted tree), parameters are the branch lengths b_1, b_2, and b_3, also measured by the expected number of substitutions per site. The clock model is a special case of the no-clock model with the constraint $b_1 = b_2$; therefore, the no-clock model reduces to the clock model when $b_1 = b_2 = t_2$ and $(b_1 + b_3)/2 = t_1$. Adapted from Yang (2014) and Rannala and Yang (2013).

The hypothesis of rate constancy can also be tasted in a likelihood framework — *the likelihood-ratio test* (LRT) — which is parallel in functionality to the relative-rate test. Given a model of nucleotide or amino acid sequence evolution and a specific tree of arbitrary size, it is possible to calculate the likelihood of generating a specific set of sequences at the terminal node of the tree. The log-likelihood value for the tree in Figure 2.1b can be computed with and without the constraint $b_1 = b_2$, ℓ_0 and ℓ_1, respectively. The first corresponds to the molecular clock hypothesis. Then $2\Delta\ell = 2(\ell_1 - \ell_0)$ is compared against X_1^2 (Muse & Weir 1992). This is a special case of the LRT of the clock (Felsenstein 1981). Under the clock model (H_0), there are $s - 1$ parameters corresponding to the ages of the $s - 1$ internal nodes on the rooted tree with s species, measured by expected number of substitutions per site. The more general no-clock model (H_1), allows every branch to have its own rate. Because time and rate are confounded, this model has $2s - 3$ parameters, corresponding to the branch lengths in the unrooted tree. The log-likelihood value for H_0 is ℓ_0 and for H_1 is ℓ_1. Then $2\Delta\ell = 2(\ell_1 - \ell_0)$ is compared against the X_1^2 distribution with $(2s - 3) - (s - 1) = s - 2$ degrees of freedom to decide whether the clock should be rejected (Felsenstein 1981; Yang & Rannala 2012; Yang 2014). In the example of Figure 2.1 the clock model (H_0) involves two parameters (t_1 and t_2 in Figure 2.1a). The more general model (H_1) does not assume the clock; thus, the parameters are the three branch lengths in the unrooted tree (b_1, b_2 and b_3 in Figure 2.1b). Then $2\Delta\ell = 2(\ell_1 - \ell_0)$ is compared against X_1^2 to decide whether the clock is rejected.

Some remarks concerning the molecular clock tests may be noted here. First, failure to reject the clock assumption does not necessarily mean that the evolutionary rate is constant over time. The null hypothesis examined by the tests described above is weaker than the assumption of a constant rate over time. For example, If the evolutionary rate has been accelerating or decelerating over time in all contemporary lineages, the tree will look clocklike, judged by distances, although the rate is not constant. Furthermore, neither the LRT nor the relative-rate tests can distinguish a variable from a constant rate within a lineage. Second, the tests examine rate differences between ingroup species and do not test whether the outgroup has a different rate from the ingroup species. For example, in the relative-rate test limited to three species only, the tests may detect a rate difference between species *a* and *b*, but not between species *c* and the two ingroup species (Figure 2.1) (Nei & Kumar 2000; Yang 2014). Finally, failure to reject the clock hypothesis may simply be

because of a lack of information in the data or lack of power of the test rather than correctness of the clock assumption. Generally, the likelihood-ratio test applied to multiple species is more powerful than the relative-rate test applied to three species only (Rannala & Yang 2013). These remarks suggest that the clock hypothesis should be accepted with caution when we estimate species divergence times from molecular data.

2.4 Statistical methods for estimating divergence times

Nowadays, modern statistical methods of divergence time estimation (Table 2.1) allow us to use both multiple fossil calibrations and sequence data, and are powerful approaches to estimate rates and divergence times under more realistic assumptions (Magallón 2004; Ho 2014). To estimate the distances from the internal nodes to the present time, both distance methods (based on estimations of pairwise distances) and likelihood methods (based on simultaneous analysis of multiple sequences on a phylogenetic tree) can be used. The choice of the substitution model is thus critical, as an oversimplified model may not correct for multiple hits adequately and can lead to underestimation of distances. The latter is generally more important for larger distances than for small ones. This non-proportional underestimation may generate systematic biases in divergence time estimates (Bromham & Penny 2003; Yang 2014)

In molecular clock dating a rooted phylogenetic tree is generally assumed to be known. However, some methods can simultaneously estimate the tree and the divergence times (Table 2.1). Phylogenetic uncertainties can severely impact divergence time estimation depending if the uncertainties affect or not the placement of fossil calibrations, and depending on the number and position of the calibration nodes. Using alternative, fully resolved phylogenetic tree topologies in a divergence dating analysis may allow us to evaluate the robustness of time estimation to uncertainties in the tree topology (Yang 2014; dos Reis et al., 2015).

Table 2.1: Likelihood and Bayesian programs that use the molecular clock to estimate evolutionary rates and timescales.

		Models of rate variation									
Program	Method	SC	LC	DC	AR	IR	Calibrations	Tip dating	Tree topology	Node-age uncertainty	Refs*
R8S	NPRS & Likelihood	•	•	•	•		Constraints	•	Fixed	Bootstrap	1
ape-chronos	Likelihood	•		•	•	•	Constraints		Fixed	Bootstrap	2
BASEML	Likelihood	•	•	•			Constraints	•	Fixed	SE/Bootstrap	3
RelTime	Likelihood		•			•	Constraints		Fixed	SE/Bootstrap	4
treePL	Likelihood	•			•		Constraints		Fixed	Bootstrap	5
BEAST	Bayesian	•	•		•	•	Priors	•	Estimated	Posterior	6
DPPDiv	Bayesian	•		•		•	Priors		Fixed	Posterior	7
MrBayes	Bayesian	•			•	•	Priors	•	Fixed	Posterior	8
MultiDivTime	Bayesian	•			•		Priors	•	Fixed	Posterior	9
MCMCTree	Bayesian	•			•	•	Priors	•	Fixed	Posterior	3
PhyloBayes	Bayesian	•			•	•	Priors		Fixed	Posterior	10
PhyTime	Bayesian				•		Priors		Fixed	Posterior	11
RevBayes	Bayesian	•	•	•	•	•	Priors	•	Estimated	Posterior	12
TreeTime	Bayesian				•	•	Priors		Estimated	Posterior	13

Models of rate variation among branches: strict clock (SC), local clock (LC), discrete clock (DC), autocorrelated relaxed clock (AR), independent rates relaxed clock (IR). *References: (1) Sanderson (2003); (2) Paradis (2013); (3) Yang (2007); (4) Tamura et al., (2012); (5) Smith and O'Meara (2012); (6) Drummond et al., (2012); (7) Heath (2012); (8) Ronquist et al., (2012b); (9) Thorne et al., (1998); (10) Lartillot et al., (2009); (11) Guindon (2013); (12) Hohna et al., (2016); (13) Himmelmann and Metzler (2009). Adapted from Ho and Duchene (2014).

Apart from potential errors of the substitution model and the tree topology, two further issues that may arise are violations of the molecular clock and uncertainties in the fossil calibrations (Rannala & Yang 2013; Ho & Duchene 2014). The likelihood and Bayesian methods of divergence time estimation are discussed below, with an emphasis on the Bayesian method. The latter can adequately account for uncertainties in fossil calibrations by specifying a prior distribution on divergence times. It can also accommodate the violation of the clock using prior models that allow substitution rate to vary across evolutionary lineages, different parts of the genome and through time. However, some likelihood models can also relax the assumption of rate homogeneity (Figure 2.2).

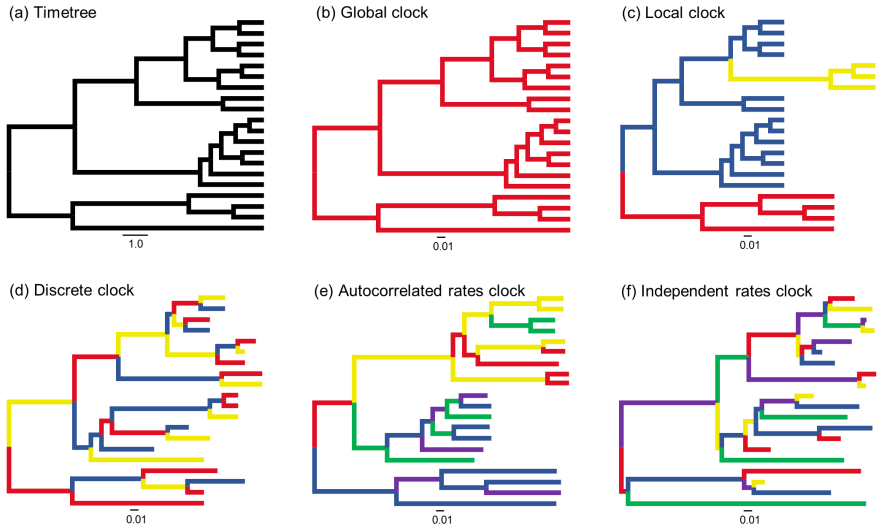


Figure 2.2: Phylogenetic trees showing differences among models of rate variation. (a) Shows a timetree with branch lengths measured in time units. The scale bar represents one time unit. The remaining subfigures show trees with branch lengths generated under different clock models: **(b)** global clock, with constant rates among branches; **(c)** local clock, with different rate in each of three group of branches; **(d)** discrete clock, with a reduced number of branch-specific rates distributed throughout the tree; **(e)** autocorrelated rates clock, with a distinct rate along each branch that is correlated with the rate along its parental branch; and **(f)** independent rates clock, with a distinct rate along each branch drawn from a selected probability distribution. In subfigures (b) to (f), scale bars represent 0.01 substitutions per site. Adapted from Ho and Duchene (2014) and Bell (2015).

2.5 Likelihood estimation of divergence times

The *global clock model* is very useful for calculating divergence times. This model is also used as the null hypothesis for testing the presence of among-lineage rate heterogeneity. As explained in subsection 2.3, a rooted tree has $s - 1$ ancestral nodes. If reliable ages determined by the fossil record are assigned to c ancestral nodes, the model involves $s - c$ parameters: the substitution rate μ and the ages of the $s - 1 - c$ non-calibrated nodes. For example, the tree illustrated in Figure 2.3 has s = 5 species, with four internal node ages: t_1, t_2, t_3, and t_4. Consider nodes of ages t_2 and t_4 fixed according to the fossil record. Then three parameters are estimated under the model: μ, t_1, and t_3. Given those rates and time parameters, each branch length — in units of expected substitutions — is simply the product of the rate and the time duration of the branch. For example, the length of the branch from nodes 2 to 3 in Figure 2.3 is $\mu(t_2 - t_3)$ (Yang 2014).

Let D represent an aligned dataset of molecular sequences and T denote the node times of the tree. With the global clock model, only a single rate R is needed. The

likelihood function $\ell(D|R,T)$, can be calculated using standard algorithms, times and rates are estimated by maximizing the likelihood function. The values of μ and T completely determine the branch lengths B on the tree. Because rates and times are confounded with molecular sequence data, $\ell(D|B) = \ell(D|R,T)$. The branch lengths $\hat{B}$ that maximize $\ell(D|B)$ are hence the ML estimates (Yang & Yoder 2003; Thorne & Kishino 2010; Yang 2014).

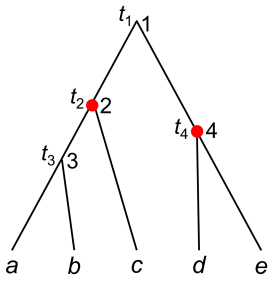


Figure 2.3: A five-species tree to explain ML and Bayesian methods of divergence time estimation. Red dotes represent fossil calibrations for nodes 2 and 4. Adapted from Yang and Yoder (2003).

The model described above assumes the molecular clock. However, to deal with violation of the clock other approaches take explicit account of among-lineage rate variation for the estimation of divergence times. For example, considering the tree of Figure 2.2, one may assign one rate for all branches on the right of the root, and another for those on the left. This approach is known as the *local clock model* (Yoder & Yang 2000; Yang & Yoder 2003) because it has clocklike evolution within prespecified regions of a tree but also allows the clock to tick at different rates among prespecified regions (Thorne & Kishino 2010). The implementation of the *local clock model* is very similar to that described above for strict molecular clock. The only difference is that under a local-clock model with k branch rates, one estimates $k - 1$ extra rate parameters. In general, it may not be clear how many categories should be modeled. Likewise, assignment of branches to individual rates may be arbitrary and because of this ad hoc assignment of where the changes have occurred, this method has not been widely used (Yang 2014; Bell 2015).

A very different approach to estimate divergence times and variable rates of evolution on a phylogenetic tree that do not rely on modeling rate changes among branches is the *non-parametric rate smoothing* (NPRS) method (Sanderson 1997). This method

uses an optimization algorithm that allow us to minimize or smooth changes in rates between adjacent branches in a phylogenetic tree, so that closely related lineages will tend to share similar rates. (Magallón 2004; Rutschmann 2006; Ho & Duchene 2014). One implementation of this approach, called *penalized likelihood* (PL) (Sanderson 2002), applies a penalty to rate changes between adjacent branches while maximizing the likelihood of the data, thus allowing estimation of both rates and times. This method includes a smoothing parameter, λ, estimated through a cross-validation (CV) procedure, which controls the magnitude of penalizing rate changes relative to the likelihood. The penalized likelihood is then maximized, allowing estimation of branch-specific rates and divergence times across the tree (Sanderson 2002). Estimates of uncertainty in node times can be obtained via parametric bootstrap (Sanderson 2003). If a probabilistic model of rate change is instead implemented (see section 2.6) there is no need for either a rate smoothing parameter or CV (Yang 2014). Another implementation of this approach, called *heuristic rate smoothing* (HRS) method (Yang 2004) differs slightly from PL, which uses a Poisson approximation to fit the branch lengths, while the HRS method uses a normal approximation of the ML estimates of branch lengths (Rutschmann 2006).

The NPRS, PL and HRS methods can deal with uncertainties in the fossil calibrations, implemented by placing minimum or maximum constraints ($t_L < t < t_U$) on the ages of calibrated nodes. However, these methods are identifiable (a property which a model must satisfy in order for precise data-dependent inference to be possible) only if at least one node age is known without error; therefore, these methods do not resolve the general problem that all fossil calibrations have some error associated with them (Rannala & Yang 2013). Although ML clock dating methods are ad hoc, and typically do not adequately accommodate different sources of uncertainty in a molecular dating analysis, these methods generally require less computation and may hence be handy for analysing massive datasets for which the Bayesian approach is still computationally prohibitive (Smith & O'Meara 2012; Ho & Duchene 2014; dos Reis et al., 2016).

2.6 Bayesian estimation of divergence times

2.6.1 Brief introduction to Bayesian inference

There are two major approaches in statistical inference, the classical or frequentist and the Bayesian. Many of the technics of frequentist statistics were developed in the early 20th century by Francis Galton & Karl Pearson (regression and correlation) and Ronald A. Fisher (analysis of variance, experimental design and likelihood). In contrast, Bayesian ideas are much older, introduced by Thomas Bayes during the 18th century and further developed by Pierre Simon Laplace and others in the 19th century (Stigler 1986; Yang 2014). In both approaches, a likelihood function $f(D|\theta)$ describes the probability of data D given the values of parameters θ. The fundamental difference of the Bayesian from the frequentist approach is that the parameter θ is treated as a random variable and thus has a distribution, while in frequentist statistics the parameter θ is an unknown constant and cannot have a distribution. Even though this difference might not seem that important, it leads to a very different statistical modeling and interpretation (Larget 2010; Yang 2014).

Inference in frequentist statistics is based on the likelihood. The value of θ that maximizes the likelihood function is an estimate of θ. Bayesian inference, by contrast, is based on the posterior distribution of θ, that is, the probability distribution that describes the uncertainty in the parameter given the data $f(\theta|D)$. To estimate the posterior distribution, we need to specify a prior distribution $f(\theta)$, which expresses the uncertainty in the parameters θ before observing any data. Then the posterior distribution, is given by Bayes' theorem (DeGroot & Schervish 2002; Larget 2010; Yang 2016)

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{f(D)}, \tag{2.1}$$

where the denominator $f(D)$ is the marginal probability of the data, averaged over all possible parameter values weighted by their prior distribution; this normalizing constant ensures that $f(\theta|D)$ is a proper statistical distribution and integrates to 1 is given by

$$f(D) = \int f(D|\theta) f(\theta) d\theta, \quad (2.2)$$

here, θ is assumed to be continuous, but if θ is discrete a sum replaces the integral. In almost all problems of pragmatic interest, it is not practical to compute $f(D)$ directly. Markov chain Monte Carlo (MCMC) methods offer a means to make Bayesian inference without the need to compute the normalizing constant (see subsection 2.6.2 for a description of MCMC).

In a frequentist approach to inference, the unknown parameter θ is fixed and generates point estimates; thus, there is no way that probabilities can be associated with it. However, confidence intervals are used to express uncertainty around the estimated values. In a Bayesian approach, the inference is the posterior distribution and generally we need to summarise the information included in the posterior into a single estimate (e.g. mean, mode and median of the posterior distribution).

The confidence interval is analogous to the credibility interval (CI) in the Bayesian framework, which is defined as $\int_{\theta_L}^{\theta_U} f(\theta|D) d\theta = 1 - \alpha$ and means that the true parameter θ is in the interval (θ_L, θ_U) with probability $1-\alpha$. A very common procedure is to build a 95% equal-tail CI using the 2.5% and 97.5% quantiles of the posterior distribution. Nevertheless, when the posterior density is multimodal or skewed this interval may include less plausible values of θ than values outside the interval. So, in the above definition one can impose the constraint that the width of the interval should include values of θ of the highest posterior density, that encompass 95% of the density mass, to form the 95% highest posterior density (HPD) interval. The HPD interval offers two advantages over the equal-tail CI: (1) any point within the HPD interval has higher density than any point outside the interval; and (2) given a probability level $1-\alpha$ the HPD interval has the smallest width. When the posterior density is unimodal and nearly symmetrical the HPD and the equal-tail CI are nearly identical (Yang 2014)

Statistical problems usually involve models with more than one unknown parameters. One might be interested in one of them or in a subgroup of them, but typically the values of the other parameters (called nuisance parameters) might be unknown. The Bayesian approach provides a way of dealing with them (Yang 2014). In case of multi-parameter problems we have a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ of parameters which we want to

make inference about. We specify a multivariate prior $f(\boldsymbol{\theta})$ and along with the likelihood function $f(D|\boldsymbol{\theta})$ we estimate the posterior using the Bayes' theorem

$$f(\boldsymbol{\theta}|D) = \frac{f(D|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(D)} = \frac{f(D|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int f(D|\boldsymbol{\theta})f(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}. \quad (2.3)$$

Note that in multivariate models, direct calculation of the normalizing constant $f(D)$ involves the calculation of a multidimensional integral which might be impossible or computationally expensive. However, the Bayesian inference has been made possible by using a MCMC algorithm which simulates from the posterior. The posterior $f(\theta|D)$ is a multivariate distribution and we can make inference about any subset of parameters by applying probability calculations. For example, the marginal posterior distribution for the parameter θ_1 can be calculated by integrating out all other parameters,

$$f(\theta_1|D) = \int f(\boldsymbol{\theta}|D)d\theta_2 \dots d\theta_p. \quad (2.4)$$

In multivariate models, there is an increased difficulty in specifying the prior distribution. The prior is a multidimensional statistical distribution that represents the scientist's best knowledge about the model and its parameters, and is specified before analysing the data. A joint prior distribution, which may accommodate the correlation structure among parameters, should be specified. However, it is a common practice to ignore the correlation among parameters, and the joint prior distribution is merely specified as the product of independent priors for the parameter. Such priors can cause problems if the posterior is sensitive to the prior.

2.6.2 Markov chain Monte Carlo

MCMC is a general computational technique for evaluating sums and integrals (Metropolis et al., 1953), especially those that emerge as expectations under complex probability distributions (Larget 2010). For most of the scientific problems it is usually necessary to build parameter-rich models. Bayesian inference usually requires the calculation of multidimensional integrals, which is not always practical. Note that all terms on the right-hand side of equation (2.3) are straightforward to calculate except the normalizing constant $f(D)$, which involves multidimensional integrals and may be too expensive to compute. Thus, MCMC is particularly suitable for Bayesian

computation. It avoids the calculation of $f(D)$ by generating a sample from the posterior via a simulation process.

MCMC methods simulate a Markov chain whose stationary distribution is the posterior distribution of the parameters of interest. The most common form of the MCMC is the Metropolis-Hasting (MH) algorithm (Metropolis et al., 1953; Hastings 1970). Let's assume that we are interested in the posterior distribution $f(\theta_1,\ldots, \theta_p|D)$. Then a sample from the posterior can be obtained by the MH algorithm. The algorithm below provides the details of a generic MH algorithm.

1. Set initial state $= \boldsymbol{\theta}^j\left(\theta_1^j, \ldots, \theta_p^j\right), j = 0.$
2. In the j+1 iteration propose a new state $\theta^* = \left(\theta_1^*, \ldots, \theta_p^*\right)$ drawn from proposal density $\mathrm{q}\left(\boldsymbol{\theta}^* \middle| \boldsymbol{\theta}^j\right)$.
3. Set the next value $\boldsymbol{\theta}^{j+1}$ in the chain $\boldsymbol{\theta}^{j+1} = \begin{cases} \boldsymbol{\theta}^*, \text{with probability } \alpha \\ \boldsymbol{\theta}^j, \text{with probability } 1-\alpha \end{cases}$, where $\alpha = \min\left\{1, \frac{f(D|\theta^*)f(\theta^*)q(\theta^j|\theta^*)}{f(D|\theta^j)f(\theta^j)q(\theta^*|\theta^j)}\right\}$.
4. Increase j. Go back to step 2 until $j = J$, where J is a fixed number of iterations.

It is worthwhile to remark that the algorithm does not involve calculation of the marginal probability as it cancels out in the calculation of the acceptance probability α (steps 2 and 3). Moreover, in step 3 it is not necessary to update all parameters at once. It is advisable to update each parameter separately or create groups of correlated parameters and update the groups one by one (Yang 2014). Updating all parameters together is complicated and might lead to poor performance of the algorithm. The states $(\theta_1^1,\ldots, \theta_p^1)$, …,$(\theta_1^j, \ldots, \theta_p^j)$ are a sample from the joint posterior $f(\theta_1, \ldots, \theta_p|D)$. The ith (i = 1, ..., p) component of each state is a sample from the marginal posterior distribution $f(\theta_i|D)$. Thus, the posterior information of any parameters can be summarised. For instance, we can approximately estimate the posterior mean for the parameter θ_i through $\mathbb{E}[\theta_i|\, D] \approx \frac{1}{J}\sum_{j=1}^{J} \theta_i^j$, or the variance and the 95% HPD interval (Larget 2010).

2.6.3 General framework of Bayesian divergence time estimation

A Bayesian MCMC algorithm was developed to estimate divergence times (Thorne et al., 1998; Kishino et al., 2001). The Bayesian method provides a framework that can simultaneously incorporate multilocus sequence information, prior information on substitution rates, prior information on rates of cladogenesis (speciation), as well as fossil calibration uncertainties, to estimate divergence times (Drummond et al., 2006; Yang & Rannala 2006; Rannala & Yang 2007). The objective of the Bayesian analysis is the estimation of posterior probability, $f(\mathbf{t},\mathbf{r},\theta|D)$, of divergence times **t**, molecular evolutionary rates **r**, and model parameters θ, given the molecular data D. This posterior is given by Bayes theorem equation (2.1)

$$f(\mathbf{t},\mathbf{r},\theta|D) = \frac{f(\theta)f(\mathbf{t}|\theta)f(\mathbf{r}|\mathbf{t},\theta)f(D|\mathbf{t},\mathbf{r},\theta)}{f(D)}, \quad (2.5)$$

Where $f(\theta)$ is the prior of model parameters; f(**t**) is the prior of times; $f(\mathbf{r}|\mathbf{t},\theta)$ is the prior of rates on the tree; and $f(D|\mathbf{t},\mathbf{r},\theta)$ is the likelihood, or the probability of the molecular data given the times, and model parameters. The marginal probability of the data, $f(D)$, is a high-dimensional integral over **t**, **r**, and θ. The MCMC algorithm is used to sample from the joint posterior. The marginal posterior of **t**,

$$f(\mathbf{t}|D) = \iint f(\mathbf{t},r,\theta|D)\mathrm{d}\mathbf{r}\,\mathrm{d}\theta, \quad (2.6)$$

can be constructed from the samples taken during the MCMC.

Below is a sketch of the MCMC algorithm implemented in the MCMCTree program in the PAML package (Rannala & Yang 2007; Yang 2007, 2014). The individual terms involved in equation (2.5) are discussed in the next several subsections.

- 1. Start with a random set of divergence times **t**, substitution rates **r**, and parameters θ.
- 2. In each iteration do the following:
  - 2a. Propose changes to the divergence times **t**.
  - 2b. Propose changes to the substitution rates for different loci.
  - 2c. Propose changes to substitution parameters θ.

- 2d. Propose a change to all times and rates, multiplying all times by a random variable *c* close to one and dividing all rates by *c*.
- 3. For every *k* iterations, sample the chain: save **t**, **r**, and θ to disk.
- 4. At the end of the run, summarise the results.

2.6.4 Calculation of the likelihood

The likelihood, $f(D|\mathbf{t},\mathbf{r},\theta)$, can be calculated under any substitution model on the sequence alignment. However, during the MCMC iteration, the likelihood function is calculated many times and is computationally expensive for large data sets. To achieve computationally efficiency an approximate method to calculate the likelihood has been developed (Thorne et al., 1998; Kishino et al., 2001; dos Reis & Yang 2011). By applying the Taylor expansion to the log-likelihood, the method calculates the gradient and the Hessian matrix of the likelihood using the maximum likelihood estimates (MLEs) of the branch lengths and the model parameters before the MCMC run. Then a transformation (e.g. square root, logarithm) is used to provide better approximation for values away from the MLEs. The approximation is efficient and allows the analysis of large datasets in realistic times (dos Reis et al., 2012; dos Reis et al., 2015). However, for a large number of partitions and large number of taxa the Bayesian method is still computationally demanding.

2.6.5 Relaxed clocks and prior model of rate drift

Bayesian molecular clock dating requires a model for the rate drift along lineages and a prior on the evolutionary rates. Two classes of Bayesian relaxed molecular clock models have been developed to deal with the among-branch rate variation, the autocorrelated-rates (AR) (Thorne et al., 1998; Kishino et al., 2001) and the independent-rates (IR) model (Drummond et al., 2006; Lepage et al., 2007; Rannala & Yang 2007), which differ in their assumptions about the nature of rate variation among branches. The relaxed clock model is incorporated in the prior $f(\mathbf{r}|\mathbf{t},\theta)$. In the AR model, the rate at each node is specified by conditioning on the rate at its ancestral node, $\log(r_A)$. For a given locus, the rate (r) at a node given the rate at the ancestral node (r_A) follows a log-normal distribution with density

$$f(r|r_{\mathrm{A}}) = \frac{1}{r\sqrt{2\pi tv}} \exp\left\{-\frac{1}{2tv}\left[\log\left(\frac{r}{r_A}\right) + \frac{vt}{2}\right]^2\right\}, 0 < r < \infty, \quad (2.7)$$

where t is the separation time between the two nodes. Specifically, the log rate (log$r$) for the current node follows a normal distribution with mean log(r_{A})−vt/2 and variance vt. This means that the logarithm of the rate drifts conforming to a Brownian motion process, while the rate itself drifts conforming to a geometric Brownian motion process. Note that the mean of the log-normal density is E(r) = r_{A} and thus the rate at the node is a value around the rate of the ancestral node. Parameter v controls how rapidly the rate drifts, which is to say how clocklike the tree is a priori. A large v means that the rates vary dramatically over time or among branches and the clock is seriously violated; on the contrary, a small v means that the clock roughly holds. The joint prior, f(**r**), of all node rates on the tree is the product of the log-normal distributions across the nodes. In the implementation of the same model by Rannala and Yang (2007) the log-normal distribution applies to the rates at the midpoint of the branches.

In the IR model, the rate for a branch is a random variable drawn from a common probability distribution such as the lognormal or the gamma. The rates evolve independently on each lineage, but the amount of rate variation has some form of evolutionary constraint which is imposed by the prior distribution on rates. For a given locus, the rate at any branch follows a log-normal distribution with density

$$f(r|\mu, \sigma^2) = \frac{1}{r\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\log\left(\frac{r}{\mu}\right) + \frac{\sigma^2}{2}\right]^2\right\}, 0 < r < \infty, \quad (2.8)$$

where μ is the mean rate for the locus and σ^2 is the variance in the log. Parameter σ^2 measures the degree of variability of the evolutionary rate, with values of σ^2 = 0.2 indicating serious clock violation.

In the IR model, the variance of the rate is independent to the time, thus the rate can undergo large shifts — depending on the value of σ^2 — including for adjacent branches. On the contrary, in the AR model the variance depends on the time and hence the model penalizes large rate variation over short time intervals but allows rate to vary nearly freely among distant clades. However, the variance increases linearly with the time and in analyses of deep phylogenies this might lead to excessively high rate shifts. Therefore, the AR model might be more suitable for the

analysis of closely-related species while the IR model for divergent species and large phylogenies. Recently, a mixed relaxed clock model that has autocorrelated- and independent-rates components has been developed (Lartillot et al., 2016). This mixed clock represents an alternative solution to the problem of choosing between autocorrelated and uncorrelated relaxed clocks. Nevertheless, it is always useful to test the robustness of time estimates to the clock model used (Drummond et al., 2006; dos Reis 2014).

2.6.6 Prior on divergence times and fossil calibrations

Fossil calibrations are the utmost source of information for resolving the distances between molecular sequences into estimates of absolute times and absolute rates in molecular clock dating analysis. There has been vast research on the use of fossil data for calibrating molecular clocks (Gandolfo et al., 2008; Benton et al., 2009; Parham et al., 2012; Warnock et al., 2015). Calibrations (Figure 2.4) are typically based on palaeontological evidence; thus a minimum constraint on the age of a clade is based on the timing of its oldest fossil representative (Ho & Duchene 2014). For the specification of maximum bounds, one could use fossils which lack major characteristics of species belonging to the clade of interest from an older geological formation, to set up a maximum constraint (Benton et al., 2009). Also, biogeographic events — such as continental split and formation of oceanic islands — can be used as maximum bounds if treated with caution (Forest 2009). The most common way to incorporate fossil calibrations into a Bayesian analysis is through the prior probability distribution placed on the age of the corresponding node. Fossil calibrations are typically used as minimum, maximum or both age bounds on node ages, and are implemented in the MCMC algorithm by not proposing new divergence times that violate such bounds. The prior for the ages of the non-calibration nodes assumes that the tree is the result of a random branching-process model (e.g., a birth-death process) (Yang & Rannala 2006; Gernhard 2008; Stadler 2009; Hohna et al., 2011).

The time prior $f(\mathbf{t})$ is constructed as

$$f(\mathbf{t}) = f(\mathbf{t}_c) f_{\mathrm{BD}}(\mathbf{t}_{\bar{c}} | \mathbf{t}_c). \quad (2.9)$$

This construction has two components: the density of ages for the nodes in the tree for which fossil calibration are available, $f(\mathbf{t}_c)$; and the conditional density, based on

the birth-death- BD process, for those nodes where no fossil calibrations are available, $f_{\mathrm{BD}}(\mathbf{t}_{\tilde{c}}|\mathbf{t}_{\mathrm{c}})$. The time prior must satisfy the constraint that ancestral nodes are older than descendent nodes. During the MCMC iteration, node ages that do not satisfy this constraint are never proposed. Effectively $f(\mathbf{t}_{\mathrm{C}})$ is truncated, and the effective prior density used by the program may differ considerably form the user specified calibration density. Therefore, it is always necessary to examine the time prior, which can be generated by MCMC iteration without sequence data. For some data sets the truncation effect may be considerable. Moreover, Bayesian divergence time programs use different procedures to construct the time prior, thus reporting the time prior is crucial for valid comparisons among different studies (Inoue et al., 2010; dos Reis et al., 2012). Chapter 4 provides a detailed analysis of fossil calibrations and construction of the time prior.

Fossil preservation biases, incorrect placement of fossils on a phylogeny or uncertainties in fossil age estimation may result in erroneous calibrations. Errors in calibrations can be spread throughout the tree, causing errors in all of the estimates of divergence times (Magallón 2004; Ho & Phillips 2009). These problems can be reduced by specifying calibrations for multiple nodes throughout the tree, which generally improve the estimates of rates and node times (Paradis 2013; Duchene et al., 2014). The manner in which the age constraint is applied, including the choice of prior density and its parameters, is usually subjective and based on a general interpretation of factors that are hard to quantify individually (Ho & Phillips 2009; Inoue et al., 2010; Donoghue & Yang 2016). Considerable effort has been spent to formalize this process to develop objective priors by proposing methods for estimating uncertainty, such as by modelling preservation and discovery (Marshall 2008; Dornburg et al., 2011; Wilkinson et al., 2011; Nowak et al., 2013; Donoghue & Yang 2016). Furthermore, studies of errors in fossil dating techniques (Gradstein et al., 2012), and morphological character evolution in fossils and modern species (Lewis 2011; dos Reis et al., 2016) may also contribute to this goal.

An alternative approach is to use a hierarchical Bayesian model in which the calibration priors can be grouped into subsets that share the same parameters. Specifically, this model samples calibrations from a mixture of exponential distributions, with the number and assignments of exponential distributions to calibration nodes being estimated (Heath 2012). Moreover, when the data set is

composed of a combination of molecular and morphological characters, fossil taxa can be included in a total evidence analysis. In such combined data sets, the fossil taxa are only represented by morphological and not molecular characters. Nonzero ages should be specified for the fossil taxa; thus, they are not contemporaneous with the extant taxa in the data set. In this way, the evolutionary relationships of the fossil taxa can be estimated, while the age of each fossil performs to constrain the times of the ancestral nodes in the tree (Stadler 2010; Pyron 2011; Ronquist et al., 2012a). Furthermore, in the fossilized birth–death (FBD) model (Heath et al., 2014), calibration of the estimates of divergence times is based on fossil occurrences. This model analyses the molecular sequence data jointly with the fossil data. It aims to describe both the distribution of fossils and the lineage divergence times within a clade based on an integrated diversification–fossilization model. The probabilities of node times are then estimated using a BD model with speciation and extinction.

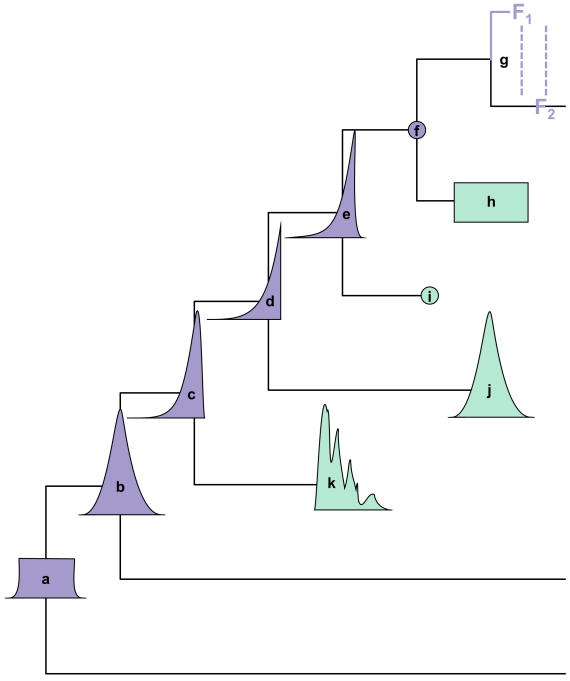


Figure 2.4: Some approaches to represent uncertainty of calibrations in a phylogenetic tree. Calibrations at internal nodes (purple): **(a)** uniform distribution with soft minimum and maximum bounds (Yang & Rannala 2006); **(b)** normal distribution; **(c)** lognormal distribution; **(d)** exponential distribution (Drummond et al., 2006); **(e)** Cauchy distribution (Inoue et al., 2010); **(f)** point value; **(g)** fossilized birth-death model (Heath et al., 2014) . Calibrations at terminal nodes (green): **(h)** uniform distribution with hard minimum and maximum bounds ; **(i)** point value; **(j)** normal distribution; **(k)** empirical calibrated radiocarbon sampler (Molak et al., 2015). Modified from Ho and Duchene (2014).

2.6.7 Uncertainties in divergence time estimates

The Bayesian approach provides a powerful framework for incorporating various sources of uncertainty in the analysis, by assigning a parametric distribution to parameters for times and rates to represent prior belief with uncertainty (Kishino et al., 2001; Drummond et al., 2006; Yang & Rannala 2006). There are four major sources of uncertainty affecting the posterior estimates of species divergence times.

The first source is the sampling errors in the estimates of branch lengths, which are due to limited sequence data. This can be reduced by sampling more molecular data. According to the finite-sites theory, part of the posterior variance of time estimates is due to limited number of sites at a locus and another to limited number of loci (dos Reis & Yang 2013; Zhu et al., 2015). The number of loci — each of which has its own relaxed clock — is the crucial factor in reducing uncertainty. Thus, increasing the number of loci seems to be more important than increasing the number of sites within each locus (Zhu et al., 2015).

The second source is the variation of substitution rates among lineages and among loci. The uncertainty due to among-branches rate variation can be reduced by sampling more loci. If it is assumed that all of the genes in an alignment have the same age — as that of the species — but differ in the pattern of evolutionary rate drift, then a long branch in a locus is more likely to be due to an accelerated evolutionary rate if this branch is short at other loci. Therefore, using multiple gene loci might be advantageous (Zhu et al., 2015).

The third source is the uncertainty in the tree topology, which can severely impact divergence time estimation depending on whether uncertainties affect the placement of fossil calibrations, and on the number and position of the calibration nodes (Ho & Phillips 2009; dos Reis et al., 2015). Thus, using alternative, fully resolved phylogenetic tree topologies in a divergence dating analysis may allow us to evaluate the robustness of time estimation to uncertainties in the tree topology (dos Reis et al., 2015).

And finally, the uncertainty in fossil calibrations, which cannot be reduced by adding more molecular data. Nevertheless, using multiple calibrations of good quality can

help to reduce the uncertainty. Fossil calibrations are crucial since they exert a significant influence on the posterior estimates due to the confounding nature of time and rate. Thus, using reliable fossil calibrations might be the most important aspect in molecular dating (Inoue et al., 2010). Moreover, the sequence data alone has an inherent limitation in time estimation and even having infinitely number of sites, there is a limit in distribution of times and rates where the errors in the time estimates do not reduce any further (Yang & Rannala 2006). Thus, the uncertainty in time estimates are usually due to fossil calibrations uncertainties rather than uncertainty in branch length estimation. Nevertheless, obtaining high quality calibration constraints is complicated, mainly because the fossil evidence is uncompleted. Objective representation of fossil information is challenging and extreme care should always be taken in the specification of calibration densities. Many suggestions have been made to improve the standard of using fossil calibrations in molecular dating, and thus reducing errors due to inappropriate calibration constraints (Parham et al., 2012; Magallón et al., 2013; Warnock et al., 2015).

3 Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution

3.1 Abstract

Angiosperms dominate plant diversity and play vital roles in terrestrial ecosystems. Through the lens of the fossil record, angiosperm diversification precipitated a Cretaceous Terrestrial Revolution (KTR) in which pollinators, herbivores and their predators underwent explosive co-diversification. Molecular dating studies instead imply that early angiosperm evolution is not documented in the fossil record. However, the reality of this mismatch remains controversial. We used powerful a Bayesian method to analyze a data set of 83 genes from 644 taxa and a suite of 52 fully-justified fossil calibrations. We explored the effect of different interpretations of the fossil record, molecular clock models and several other sources of uncertainty on angiosperm divergence time estimates. We found that fossil calibration uncertainty has a strong effect on estimated divergence times. Using Bayesian model selection, we demonstrate that the IR model fits the tracheophyte data better. Controlling for different factors of uncertainty does not bring molecular and palaeobotanical estimates into agreement. We reject a post-Jurassic origin of angiosperms, supporting the notion of a cryptic early history of angiosperms. These conclusions remain compatible with palaeobotanical evidence and with the more general hypothesis of a KTR in the sense that diversification of the major groups of angiosperms occurred later within the Cretaceous, alongside the important diversification of other lineages.

Keywords: Bayesian analysis, divergence time, fossil record, angiosperms, Cretaceous Terrestrial Revolution.

3.2 Introduction

The angiosperms represent one of the largest branches of the tree of life. They dominate extant plant diversity, occupy almost every habitat on Earth, and are one of the principal components of modern biota playing crucial roles in terrestrial ecosystems(Augusto et al., 2014; Cascales-Minana et al., 2016). Angiosperms rose to ecological dominance in the KTR, when their apparently explosive radiation is believed to have underpinned the diversification of lineages that are main contributors to extant terrestrial environments such as birds, insects, mammals, and seed-free land plants foreshadowing modern terrestrial biodiversity (Dilcher 2000; Benton 2010; Meredith et al., 2011; Cardinal & Danforth 2013; Augusto et al., 2014; Cascales-Minana et al., 2016). However, these hypotheses of co-diversification rest largely on the perceived coincidence in the radiation of angiosperms and the renewal of trophic networks in terrestrial ecosystems. This is evidenced, not least, by the fossil record of tricolpate pollen in the Barremian, slightly younger Aptian floral assemblages, followed by an explosive increase in diversity in the middle and late Cretaceous (Doyle 2008; Clarke et al., 2011; Magallón et al., 2015; Herendeen et al., 2017). Some interpret this evidence literally to reflect an explosive radiation from a Cretaceous crown ancestor, with the earliest macrofossil record of an unambiguous angiosperm (Friis et al., 2000b; Sun et al., 2002) dating back only to the mid-Early Cretaceous (Hickey 1997; Dilcher 2000; Benton 2010; Friis et al., 2010; Meredith et al., 2011; Doyle 2012; Gomez et al., 2015; Cascales-Minana et al., 2016; Herendeen et al., 2017). In stark contrast, molecular timescales for angiosperm evolution have invariably concluded that crown angiosperms diverged as much as 100 million years (Myr) earlier than the KTR (e.g. Bell et al., 2005; Bell et al., 2010; Magallón 2010; Smith et al., 2010; Clarke et al., 2011; Magallón et al., 2013; Magallón 2014; Zanne et al., 2014; Zeng et al., 2014; Beaulieu et al., 2015; Foster et al., 2016; Murat et al., 2017) — unless they have been forced to fit within the age of the oldest fossil angiosperms (Magallón & Castillo 2009; Magallón et al., 2015) — (Table 3.1), implying a long cryptic evolutionary history unrepresented in the fossil record. This may be because early angiosperms were not ecologically significant, or were living in environments where fossilization was unlikely (Raven & Axelrod 1974; Feild et al., 2009; Friedman 2009; Smith et al., 2010; Doyle 2012). Or it may be that molecular clock estimates are just unrealistically old, perhaps an artifact of their failure to accommodate dramatic accelerations that may have been associated with an

explosive diversification of angiosperms (Magallón 2010; Smith et al., 2010; Beaulieu et al., 2015; Brown & Smith 2017).

Moreover, the timescale of angiosperm diversification varies broadly among different molecular analyses (Table 3.1). This is not surprising given that transforming molecular distances (the branch lengths on a phylogeny) into geological divergence times is challenging (dos Reis & Yang 2013). Certainly, there are a number of methodological variables in previous molecular analyses, that are known to affect the accuracy and precision of divergence time estimates. Foremost among these is the cursory approach taken in establishing fossil calibrations, which have been shown to contribute the greatest source of uncertainty associated with molecular clock analyses (Sauquet et al., 2012; Warnock et al., 2012; dos Reis & Yang 2013; Magallón et al., 2013). Hence, a suite of best practices has been established for formulating fossil calibrations (Parham et al., 2012), but this have not generally been applied to angiosperms. Foster et al. (2016) have highlighted the particular challenge of dating angiosperm divergence accurately using the low taxon sampling common to theirs and other studies (e.g. Bell et al., 2005; Bell et al., 2010; Magallón 2010; Smith et al., 2010; Clarke et al., 2011; Magallón et al., 2013; Magallón 2014; Zeng et al., 2014; Beaulieu et al., 2015; Foster et al., 2016; Murat et al., 2017). Some previous analyses are also limited by either insufficient outgroup lineages (e.g. Bell et al., 2005; Bell et al., 2010; Zeng et al., 2014; Magallón et al., 2015; Foster et al., 2016), very limited sequence data (e.g. Bell et al., 2005; Magallón & Castillo 2009; Bell et al., 2010; Magallón 2010; Smith et al., 2010; Clarke et al., 2011; Magallón et al., 2013; Magallón 2014; Beaulieu et al., 2015; Magallón et al., 2015), and usually a combination thereof. All of these methodological variables are known to affect the accuracy and precision of divergence time analyses (dos Reis et al., 2016).

Table 3.1: Overview of estimates of divergence times for selected major groups of angiosperms for some selected analyses from previous studies

Study	Data / Analysis	Clade (crown group)							
		Angiosperms	Magnoliids	Monocots	Eudicots	Superrosids	Rosids	Superasterids	Asterids
Bell et al. (2005)	Loci: 2-plastid, 1-mt, 1-nuc. Taxa: 71. Calib: 5. / BRC	140 – 180 Ma	—	99 – 133 Ma	93 – 125 Ma	—	—	—	—
	Loci: 2-plastid, 1-mt, 1-nuc. Taxa: 71. Calib: 5. / PL	155 – 198 Ma	—	123 – 126 Ma	—	—	—	—	—
Magallón & Castillo (2009)	Loci: 3-plastid. Taxa: 256. Calib: 13. / PL	130 – 242 Ma	—	—	—	—	—	—	—
Bell at al. (2010)	Loci: 2-plastid, 1-nuc. Taxa: 567. Calib: 36a. / IR	141 – 154 Ma	121 – 130 Ma	—	123 – 134 Ma	111 – 121 Ma	97 – 105 Ma	113 – 132 Ma	98 – 111 Ma
	Loci: 2-plastid, 1-nuc. Taxa: 567. Calib: 36b. / IR	167 – 199 Ma	108 – 138 Ma	—	123 – 139 Ma	111 – 135 Ma	97 – 132 Ma	113 – 131 Ma	98 – 119 Ma
Smith et al. (2010)	Loci: 2-plastid, 1-nuc. Taxa: 154. Calib: 33. / IR	182 – 257 Ma	136 – 181 Ma	139 – 167 Ma	128 – 147 Ma	—	—	—	—
	Loci: 2-plastid, 1-nuc. Taxa: 154. Calib: 32. / IR	193 – 270 Ma	138 – 198 Ma	141 – 191 Ma	138 – 172 Ma	—	—	—	—
Clarke et al. (2011)	Loci: 7-plastid. Taxa: 18. Calib: 17. / IR	175 – 240 Ma	—	—	83 – 115 Ma	—	—	—	—
Magallón et al. (2013)	Loci: 5-plastid. Taxa: 80. Calib: 28. / IR	162 – 210 Ma	131 – 155 Ma	125 – 145 Ma	120 – 129 Ma	—	—	—	—
Magallón (2014)	Loci: 5-plastid. Taxa: 81. Calib: 27. / IR	162 – 210	—	—	—	—	—	—	—
Zanne et al. (2014)	Loci: 11-plastid, 4-mt, 2-nuc. Taxa: 32,223. Calib: 39. / PL	243 Ma	147 Ma	171 Ma	137 Ma	118 Ma	117 Ma	117Ma	108 Ma
Zeng et al (2014)	Loci: 59-nuc. Taxa: 61. Calib: 2. / IR	286 – 246 Ma	122 – 150 Ma	127 – 149 Ma	115 – 126 Ma	—	—	—	—
Magallón et al. (2015)	Loci: 3-plastid, 2-nuc. Taxa: 798. Calib: 137. / IR	139.4 Ma	130 – 134 Ma	132 – 135 Ma	130 – 133 Ma	119 – 125 Ma	115 – 123 Ma	120 – 126 Ma	110 – 119 Ma
Beaulieu et al. (2015)	Loci: 3-plastid, 1-nuc. Taxa:125. Calib: 24. / IR	210 – 253 Ma	160 – 195 Ma	149 – 181 Ma	142 – 170 Ma	124 – 144 Ma	113 – 136 Ma	120 – 143 Ma	99 – 119 Ma
Foster et al. (2016)	Loci: 76-plastid. Taxa:195. Calib: 37. / IR	192 – 251 Ma	130 – 171 Ma	141 – 176 Ma	136 – 154 Ma	123 – 135 Ma	118 – 131 Ma	107 – 126 Ma	108 – 124 Ma

Murat et al. (2017)	Loci: 1,175. Taxa:37. Calib: 2. / IR	190 – 238 Ma	—	—	87 – 109 Ma	—	—	—	—
This study (composite)	Loci: 77-plastid, 4-mt, 2-nuc. Taxa: 644. Calib: 52. / IR	149 – 256 Ma	128 – 190 Ma	123 – 181 Ma	129 – 188 Ma	118 – 162 Ma	117 – 160 Ma	118 – 164 Ma	107 – 146 Ma

Notes: BRC: Bayesian relaxed clock (Multidivetime); PL: Penalized likelihood; AR: autocorrelated rates model; IR: independent rates model; SC strict clock model; Calib: calibration points; composite: 95% high posterior density credibility interval (HPD CI) is a composite of the 95% HPD credibility intervals across all calibration strategies, except calibration strategy B (SB). See original works for further information on time estimates.

In an attempt to explore the impact of these variables on the mismatch between molecular clock estimates and fossil evidence for the origin and diversification of angiosperms, we compiled a molecular sequence dataset of nucleotide and amino acid data from 83 plastid, mitochondrial and nuclear genes, from 644 taxa (Soltis et al., 2011; Ruhfel et al., 2014). This is the largest taxon-locus dataset employed to address the timescale of angiosperm diversification, encompassing the diversity of angiosperms as well as seed plant, fern, and lycophyte outgroups, simultaneously addressing concerns of taxon and locus diversity, as well as outgroup inclusion. We used these data to estimate tracheophyte interrelationships by Maximum Likelihood (ML). We employed 52 fossil calibrations — all of which achieve the expectations of best practice (Parham et al., 2012). We combined this in a Bayesian relaxed clock divergence time analysis. The Bayesian approach used here (Rannala & Yang 2007; dos Reis & Yang 2011) integrates over the uncertainty in rate variation along the phylogeny, allowing the analysis a large taxon-locus dataset in a larger combination of analysis set-ups that what has been possible before to explore the impact of different sources of uncertainty on the timescale of angiosperm diversification. We employed five calibration strategies that accommodate different interpretation of the fossil record and show that these have a strong impact. We also perform sensitivity tests to explore and control for sources of uncertainty including data partitioning, parameter choice in priors for rates and times, relaxed molecular clocks, and the effect of outgroup sampling. As an unexpected result, the difference in age estimates for most shallow and some deep nodes of the phylogeny differ considerably when comparing the molecular rate models. To resolve this, we developed a new method, implemented in MCMCTree, that uses Bayesian model selection to discriminate among clock models. This analysis finds that the IR model (Rannala & Yang 2007) fits the tracheophyte data better, and time estimates under this model are preferred, diminishing uncertainty in our evolutionary timescale. The sensitivity tests also show that the effect of data partitioning has a significant impact on divergence time estimates. These experiments allow us to establish a timescale for angiosperm evolution that integrates sources of uncertainty that cannot be constrained. The resulting timescale allows us to reject, with confidence, a post-Jurassic origin of crown-angiosperms, though estimates for the origin of the most diverse clades of angiosperms are in closer agreement with the fossil evidence than previous unconstrained molecular clock analyses.

3.3 Material and methods

3.3.1 Molecular data assembly

We assembled a dataset comprising 83 genes from 644 taxa (632 angiosperms, 8 gymnosperms, 2 monilophytes and 2 lycophytes) from three sources. First, sequences for 16 genes (10 plastid, 4 mitochondrial and 2 nuclear) from 640 taxa (632 angiosperms and 8 gymnosperms, 2 ferns and 2 lycophytes that serve as hierarchically-nested outgroups) were retrieved from GenBank using the accession numbers from Soltis et al. (2011). As many gene sequences in the alignment of Soltis et al. (2011) were partial sequences or a mixture of coding and non-coding segments (introns or spacers), we cleaned and curated their list of GenBank accession numbers and retrieved the sequences again. CDS sequences for each coding gene as well as partial or complete sequences for nuclear rRNA genes were retrieved. Each gene was realigned using the MAFFT algorithm (Katoh & Standley 2013) implemented in TranslatorX (Abascal et al., 2010) and curated. This process did not recover the original alignments of Soltis et al. (2011) and extra species and gene sequences previously missing or incomplete were added to the data set. Second, sequences for 78 plastid genes from 110 taxa (2 monilophytes, 2 lycophytes, 6 gymnosperms and 100 angiosperms) were taken from Ruhfel et al. (2014). Eleven genes in the data set were found to be also in the dataset of Soltis et al. (2011), and were removed. Third, sequences for additional 16 genes (10 plastid, 4 mitochondrial and 2 nuclear) from 2 monilophytes and 2 lycophytes were obtained from GenBank, aligned using MAFFT. Gene alignments from all three sources were combined into one dataset using SeaView (Gouy et al., 2010).

For each gene, a phylogenetic tree was constructed by ML using RAxML 7.7.8 (Stamatakis et al., 2005) (Table B.1). Sequences with unusually long external branches (that accounted for more than 30% or the total tree length) were removed *(nad5* for *Selaginella* and *rps4* for *Huperzia*). GenBank accession numbers for all sequences are available on Figshare. The final alignment includes 83 genes and has 75030 base pairs (bp) with 71.4 % missing data. This was divided into five partitions: (1) 1st and 2nd codon positions for plastid genes; (2) 3rd positions for plastid genes; (3) 1st and 2nd codon positions for mitochondrial genes; (4) 3rd positions for mitochondrial genes; and (5) nuclear RNA genes. However this large amount of

missing data did not seem to be an impediment to this combined approach (Roure et al., 2013; Zheng & Wiens 2016) , the broad phylogenetic relationships were very similar to those from analysing 81 taxa (36% missing data) or 48 taxa (26% missing data). Some basic information about those five partitions obtained using RAxML such as the tree length and tree topology is given in Table B.2 and Figures B.4-B.6 (molecular datasets and GenBank accession numbers available on Figshare: https://figshare.com/s/404b70bc39656c2cf57e)

3.3.2 Tree topology

The final alignment, with five partitions as described above, was used to estimate the ML tree using RAxML, under the GTR+ Γ model with 100 bootstrap replicates. The model assumes independent substitution parameters, with joint branch length optimization. The ML tree (Figure 3.1 and Figure B.3) was used for subsequent molecular clock dating analyses.

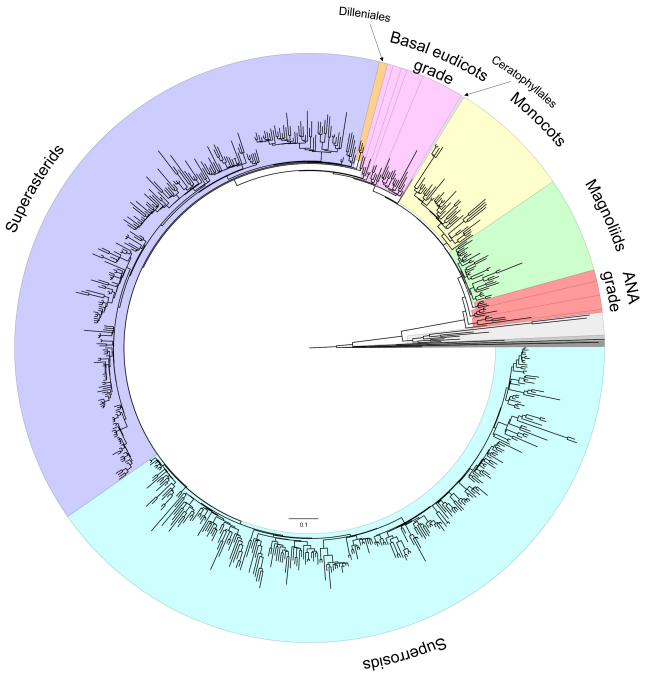


Figure 3.1: RAxML tree estimated from the 83 genes and 644 taxa of tracheophytes. The major angiosperm lineages and grades are highlighted: ANA grade (red), magnoliids (green), monocots (yellow), Ceratophyllales (pale blue), basal eudicots grade (pink), Dilleniales (orange), superasterids (purple) and superrosids (blue). Species names and bootstrap support values and are indicated in Figure B.3.

3.3.3 Fossil calibrations

Bayesian clock dating was conducted using the MCMCTree program from the PAML4.8 package (Yang 2007) incorporating soft-bound fossil calibrations on nodes on the tree (Yang & Rannala 2006). Node age constraints are based on fossils that have been placed in groups with unequivocal synapomorphies and/or through phylogenetic analysis (Figure 3.2, Table B.3 and Supplemental Experimental Procedures). The inclusion of hierarchically-nested outgroups allows us to take advantage of the effects of truncation in the construction of the joint time prior, which serves to preclude phylogenetically incompatible clade ages (i.e. ancestral nodes younger than descendants) from being proposed simultaneously to the MCMC (Inoue et al., 2010). In this way, the conservative maximum constraint on the age of the angiosperm total group will be diminished because of temporal overlap with the specified time prior on the spermatophyte, euphyllophyte, and tracheophyte clades.

We employed five calibration strategies to accommodate different interpretations of the fossil record. In the first, strategy A (SA), the eleven calibrations for which soft maximum constraints are available (Fig. 2 and Table S3) are modelled using a prior probability of 94% for a uniform distribution bounded by the minimum and maximum fossil constraints B(t_L, t_U, p_L, p_U),, and a 1% power decay distribution on the minimum constraint (p_L = 0.01), and a 5% exponential decay on the maximum constraint (p_U = 0.05). The remaining 41 calibrations nodes have minimum bounds only (Fig. 2 and Table S3), specified using a truncated Cauchy distribution L(t_L, p, c, p_L), where p determines how far from the bound the mode of the distribution is, c determines how sharply the distribution decays to zero and p_L is the left tail probability (Inoue et al., 2010). We used p = 0.1, c = 0.1 and p_L = 0.01. In the second, strategy B (SB), the 41 node calibrations with minimum bound inherit the maximum bound from the youngest ancestor which has a maximum bound so that each of the 52 calibrations has a pair of minimum and maximum bounds. The prior probability of clade age was established by a uniform distribution between minimum and maximum bounds reflecting agnosticism about the true time of divergence between these bounds. Again, we used p_L = 0.01 and p_U = 0.05. The remaining three calibration strategies C to E (SC-E) follow the first (SA), but implement different calibration densities for the crown of angiosperms (node 648 in the tree of Fig. S2) and mesangiosperms (node 451 of Fig. S2). In SC and SD, we used the truncated Cauchy distribution with either a medium

tail (*c* = 0.01) (SC) or a short tail (SD) (*c* = 0.005) extending back in time, reflecting a view that the fossil minimum constraints are increasingly poor approximations of clade age. For completeness, to explore the impact of accepting the conventional palaeobotanical interpretation of a Cretaceous origin of crown-angiosperms (e.g Herendeen et al., 2017), analysis SE used an optimistic maximum (139.4 Ma) soft bound for crown angiosperms and crown mesangiosperms based on an estimate of Magallón et al. (2015). The time unit was set to 100 Myr (phylogenetic trees in Newick format with fossil calibrations available on Figshare: https://figshare.com/s/404b70bc39656c2cf57e).

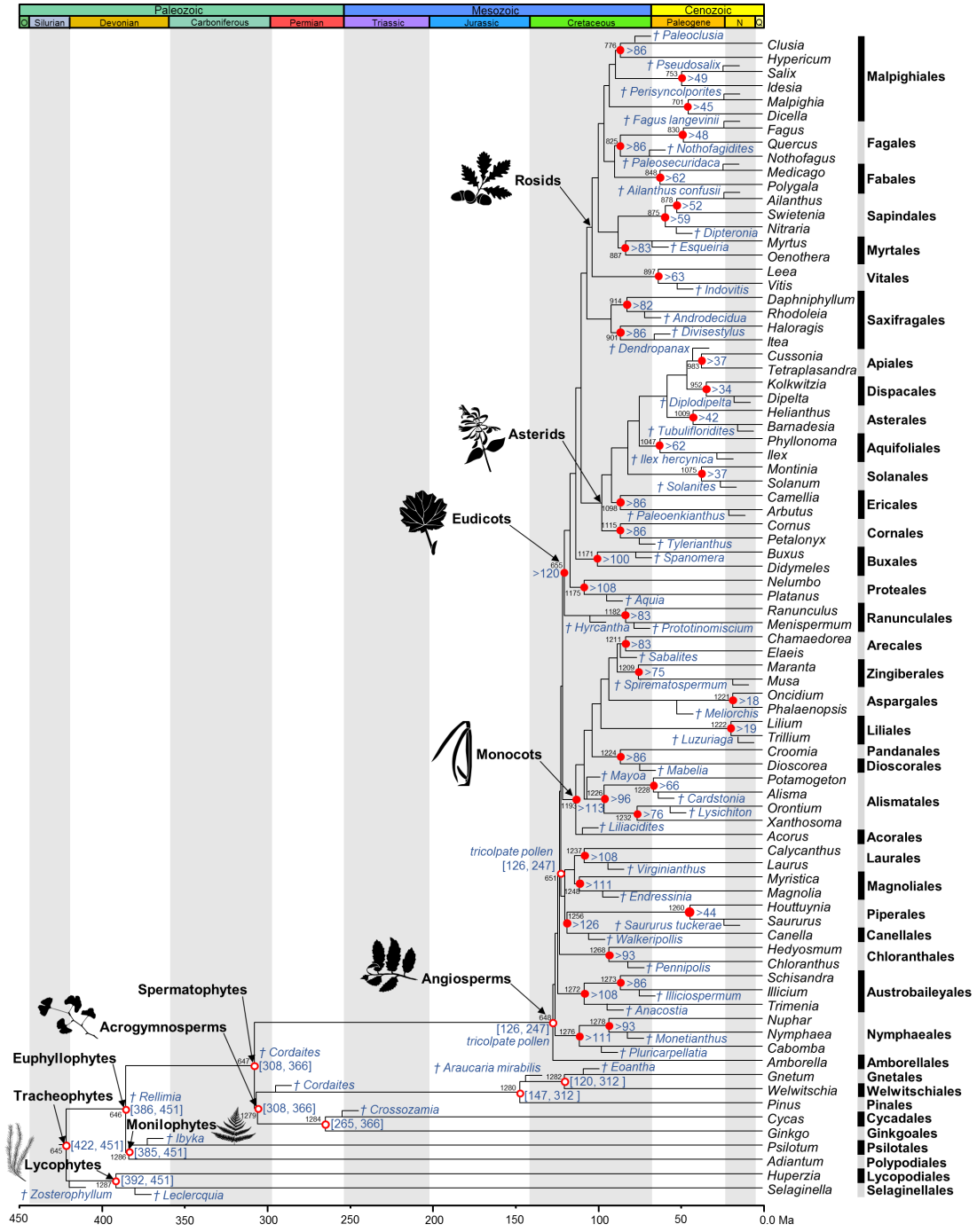


Figure 3.2: Summary tree of tracheophytes showing fossil calibrations. Calibrations are represented for 52 nodes, consisting of (>) soft minimum or both ([min, max]) soft minimum and soft maximum. Calibrated nodes are numbered as in Figure B.2. Justifications for these minima and maxima are provided in the Supplemental Experimental Procedures and an overview in Table B.4. The dagger symbol shows a species which is extinct. The tree has been scaled to time on the basis of the minimum constraints.

3.3.4 Bayesian divergence time estimation

To examine the robustness of the posterior time estimates several analyses were performed by changing prior assumptions and parameters settings. These include data partitioning, calibration strategies, parameter choice for priors for rates and times, birth-death process parameters and exclusion of distantly related outgroups with very long branches.

Our dating analyses used three of the five partitions described earlier, with the two partitions for third codon positions (in plastid and mitochondrial genes) excluded. The alignment had 51792 bp, with 70.5% missing data. Our "standard" analysis (SA-IR-3P) uses calibration strategy A, independent-rates (IR) model (Thorne et al., 1998; Rannala & Yang 2007) and HKY85+Γ5 substitution model (Yang 2007) , with three partitions. The three partitions were (1) 1st and 2nd codon positions for plastid genes, (2) 1st and 2nd codon positions for mitochondrial genes, and (3) nuclear RNA genes, as described above. In the IR model, the rate for any branch is a random variable from a lognormal density LN(μ, $\sigma 2$), where μ is the mean of the rate and $\sigma 2$ is the variance of the log rate. A gamma prior G(2, 50) was specified for μ, with mean 0.04 substitutions per site per 100 Myr or 4 x 10^{-10} substitutions per site per year (s/s/y). This is based on rough estimates of substitution rates obtained by fitting a strict molecular clock to the sequence data, using a point calibration (vascular plants, 438 Ma) on the root. A gamma prior G(2, 4) was assigned for σ^2 , with mean 0.5. The prior on times was constructed using fossil calibration densities combined with the birth-death-sampling process, which specifies the distribution of the ages of non-calibrated nodes (Yang & Rannala 2006). The parameter values $\lambda = \mu = 1$ and $p = 0$ specified a uniform kernel.

We conducted ten additional analyses that are variations of the standard analysis to examine the robustness of the posterior time estimates. We examined the truncation effect among the calibrated nodes, by generating the joint prior of times by running the MCMC without data. We used the four alternative calibration strategies to assess the impact of calibration strategy, resulting in Analyses SB-IR-3P, SC-IR-3P, SD-IR-3P, and SE-IR-3P. To assess the effect of the number of partitions, we set up two analyses. In Analysis SA-IR-1P, the three partitions were concatenated and treated as a single partition, and in Analysis SA-IR-MP, a mixed alignment, divided into plastid

proteins, mitochondrial proteins and nuclear RNA genes was used. To assess the impact of the birth-death-sampling prior, the parameters of the birth-death model were altered such that the kernel has an L shape (λ = 1, μ = 4, and ρ = 0.1), giving a tree with long internal branches (Analysis SA-IR-3P-BD1), or an inverted L shape (λ = 4, μ = 1, and ρ = 0.0001), giving a tree with long terminal branches (Analysis SA-IR-3P-BD2). To assess the effect of the rate model, Analysis SA-AR-3P was conducted under the autocorrelated rates (AR) model (Rannala & Yang 2007). Finally, to explore the effect of excluding distantly related outgroups, lycophytes and monilophytes were removed from the alignment (Analysis SA-IR-3P-EP). In this analysis, we used a gamma prior G(2, 60) for μ with mean 0.03 substitutions per site per 100 Myr or 3 x 10-10 s/s/y, based on a rough substitution rate estimate obtained by fitting a strict molecular clock to the sequence data, using a point calibration (seed plants, 337 Ma) on the root.

The likelihood (or the probability of the sequence alignment given the tree and branch lengths) was calculated using the approximate method (Thorne et al., 1998; dos Reis & Yang 2011), using the SQRT transformation (dos Reis & Yang 2011). ML estimates of branch lengths and the Hessian matrix were calculate using the programs BASEML and CODEML. We used the HKY85+Γ5 model for nucleotide alignments, the cpREV64 substitution model for plastid proteins and the WAG model for the mitochondrial proteins. For each analysis, the MCMC was run for ~5.5 million iterations after a burnin of 250,000 iterations. The chain was sampled every 80 iterations until ~70,000 samples were collected. Each analysis was done at least twice, and consistency between runs was used as a major check on MCMC convergence. We also compared the posterior mean times and plotted the time series traces using the MCMC samples. The resulting posterior distribution was summarised as the posterior means and 95% equal-tail credibility intervals (CIs) for divergence times.

3.4 Results

3.4.1 Topology estimation and the effect of fossil calibration uncertainty

Employing an unprecedented number of taxa and loci, we recovered a topology in which these deep-level relationships among angiosperms are resolved with

confidence and most branches are supported with bootstrap value of 100% (Figure 3.1 and Figure B.3).

To explore the robustness of angiosperm divergence time estimates to calibration choice, we employed five calibration strategies that share the same palaeontological constraints (Fig. 3.2, Table B.3 and Supplemental Experimental Procedures) but differ in their interpretation of this evidence, expressed as different statistical distributions (Fig. B.1). We used the program MCMCTree to obtain posterior time estimates under these five calibration strategies on the fixed tree topology derived from our phylogenetic analysis (Figs. 3.1 and B.3). The IR model was used to specify the prior of evolutionary rates on branches on the tree topology. The 83 genes were subdivided and analysed as 3 partitions (3P) and under the HKY85+Γ5 substitution model, with third codon positions excluded from all analyses. In all cases, we first ran the analyses without sequence data to calculate the marginal priors on node ages, and assess the impact of truncation (Inoue et al., 2010; Warnock et al., 2015).

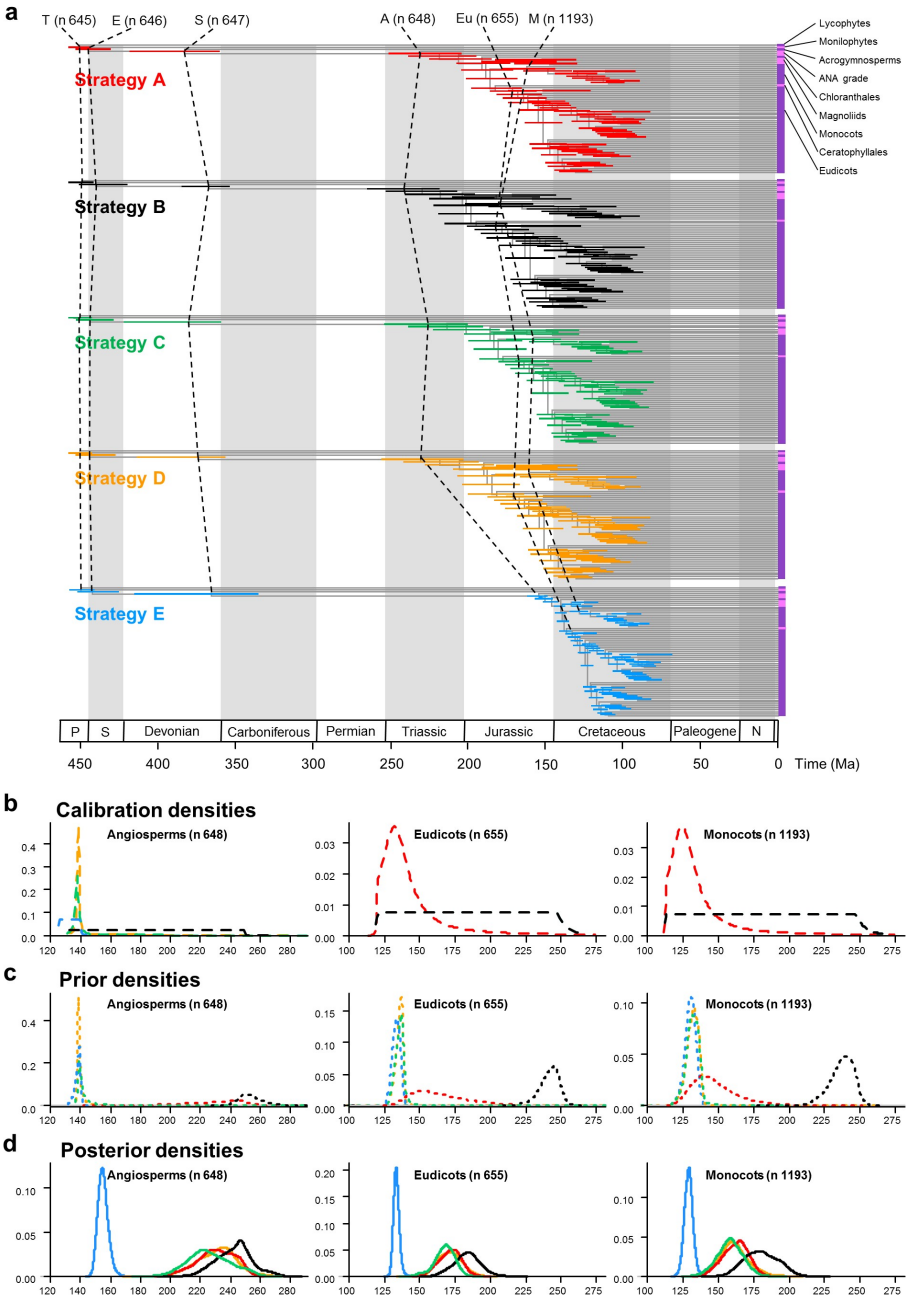


Figure 3.3: The effect of calibrations on posterior divergence time estimates of major groups of tracheophytes and angiosperms (a) Summary chronogram for tracheophytes (including 2 lycophytes, 2 monilophytes, 8 gymnosperms and 64 orders of angiosperms) with terminals collapsed to represent angiosperm orders showing divergence time estimates. Nodes are drawn at the posterior means obtained and horizontal bars represent 95% HPD CI. Estimates were obtained using the HKY85+Γ5 substitution model, IR, with the 83 genes subdivided into three partitions: (1) 1st and 2nd codon positions for plastid genes; (2) 1st and 2nd codon positions for mitochondrial genes; and (3) nuclear RNA genes. Five nodes are connected across the analyses to facilitate comparison: tracheophytes (n 645), seed plants (n 647), angiosperms (n 648), eudicots (n 655) and monocots (n 1193). **(b-d)** Calibration, prior, and posterior densities for 3 angiosperm nodes in the tracheophytes phylogeny. Colouring relates to the calibration strategy as in (a). The phylogeny with clade names is provided in Figure 3.5. Nodes in parenthesis are numbered as in Figure B.2.

The results of these analyses demonstrate that calibration strategy has a strong impact on estimated divergence times (Figure 3.3a, Figures 3.4g and 4j, Table 3.2 and Figure B.1). Estimates based on SA indicate that crown angiosperms originated 255-206 Ma, crown-eudicots 186-156 Ma, and crown-monocots 179-144 Ma (Table 3.2 and Figure B.2). Using shorter tail calibration densities on the key nodes of crown angiosperms and crown mesangioperms (SC, SD) had no significant impact on the resulting posterior time estimates (Figure 3.3d, Figures 3.4h and 3.4i, and Table 3.2). In contrast, calibration strategy SB produced older estimates and larger intervals than all the other calibration strategies (crown angiosperms 266-219 Ma, crown eudicots 201-164 Ma and crown monocots 203-127 Ma; Figure 3.3d, Table 3.2 and Figure B.1). This occurs because this calibration strategy is uninformative on the timing of divergence between minimum and maximum constraints and the effect of truncation in the construction of the joint time prior results in effective priors on node ages that places the majority of the probability mass near the maximum age bound (Figure 3.3d, Table 3.2 and Figure B.1). In effect, the fossil minima are considered a poor approximation of clade age. This is particularly apparent in the marginal priors (and posteriors) for crown clades of angiosperms, mesangiosperms, monocots, eudicots (Figures 3.3c and 3d, and Figure B.1), Alismatales, Laurales and stem Canellales. Calibration strategy SE considered whether molecular estimates could be forced into agreement with fossil evidence, employing an unrealistically optimistic 139.4 Ma maximum constraint on the age of crown-angiosperms. Unsurprisingly, this yielded significantly younger and more precise time estimates for crown clades of angiosperms (162-149 Ma), eudicots (137-129Ma), and monocots (135-123 Ma), along with many other clades (Figure 3.3, Figure 3.4j, Table 3.2 and Figure B.1). Nonetheless, the inferred age of crown-angiosperms remains significantly older than the earliest unequivocal fossil evidence (~126 Ma).

Table 3.2: The 95% HPD limits of posterior divergence times, in millions of years before the present, for selected nodes in the vascular plant tree under the 5 calibration strategies.

		SA-IR-3P		SB-IR-3P		SC-IR-3P		SD-IR-3P		SE-IR-3P		Composite	
Node	Clade	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
645*	Tracheophytes	444	458	444	460	443	458	443	458	442	458	442	458
646*	Euphylophytes	433	455	424	455	430	454	430	455	427	453	427	455
647*	Spermatophytes	360	418	356	387	359	422	356	413	340	418	340	422
648*	Angiosperms	206	253	219	266	201	255	203	256	149	162	149	256
651*	Mesangiosperms	173	208	183	225	168	203	171	210	138	145	138	210
655*	Eudicots	156	186	164	201	152	181	155	188	129	137	129	188
661	Superrosids	136	160	141	173	134	157	136	162	118	126	118	162
662	Rosids	135	159	140	172	133	155	135	160	117	125	117	160
668	**Malpighiales**	113	132	115	144	111	130	113	133	100	110	100	133
701*	Stem-*Malphighia*	44	51	35	53	44	51	45	51	44	51	44	51
753*	Stem-*Salix* plus *Populus*	48	55	38	55	48	55	48	55	48	54	48	55
776*	Stem-Clusiaceae	88	107	88	117	88	105	88	107	85	93	85	107
796	**Oxidales**	100	134	105	144	99	132	100	134	86	110	86	134
809	**Celastrales**	89	123	93	133	88	122	89	123	78	104	78	123
825*	**Fagales**	86	108	86	123	86	108	86	108	85	98	85	108
830*	Fagaceae	48	64	46	78	48	64	48	64	48	62	48	64
832	**Cucurbitales**	76	113	80	121	75	111	77	113	65	96	65	113
837	**Rosales**	92	122	97	132	91	120	91	122	81	104	81	122
847	**Fabales**	88	124	98	136	87	122	89	124	77	103	77	124
848*	Stem-*Polygalaceae*	79	116	90	128	78	114	80	117	69	96	69	117
855	**Zygophyllales**	50	112	52	118	49	111	49	112	43	95	43	112
862	**Brassicales**	81	107	84	115	78	104	81	107	69	90	69	107
869	**Malvales**	71	102	76	110	71	101	72	103	61	88	61	103
874	**Huertales**	28	96	30	103	28	94	29	97	23	81	23	97
875*	**Sapindales**	66	96	71	107	66	95	66	96	64	84	64	96
878*	Stem-*Alianthus* plus *Citrus*	52	62	47	69	52	62	52	62	52	60	52	62
881	**Crossosomatales**	90	132	97	144	89	130	91	133	81	111	81	133
887*	**Myrtales**	84	114	89	130	84	112	85	115	83	99	83	115
895	**Geraniales**	14	48	8	50	12	52	7	60	9	41	9	60
897*	**Vitales**	63	80	58	105	63	80	62	80	62	79	62	63
898	**Saxifragales**	97	126	100	142	97	125	97	127	94	111	94	127
901*	Saxifragales core	87	113	89	128	88	112	88	114	86	100	86	114
914*	Stem-Hamamelidaceae	81	97	75	104	81	97	81	97	81	97	81	97
921	Superasterids	135	159	140	171	133	156	135	161	116	126	116	161
924	Asterids	121	145	127	156	120	143	121	146	107	118	107	146
933	**Dispacales**	75	100	77	106	74	98	74	99	67	87	67	100
952*	Stem-Dipelta	34	40	13	36	34	40	34	40	34	40	34	40
962	**Paracryphiales**	50	100	52	105	49	99	49	99	43	88	43	100
964	**Apiales**	83	109	85	115	82	108	83	109	74	93	74	109
983*	Araliaceae core	37	42	21	38	37	42	37	42	37	42	37	42
993	**Bruniales**	76	111	78	117	75	110	76	112	67	95	67	112
996	**Escalloniales**	76	112	79	118	75	110	77	112	66	96	66	112
1002	**Asterales**	87	109	90	116	86	108	87	109	77	93	77	109
1009*	Stem-Asteraceae minus *Barnadesia*	45	67	51	74	45	66	45	67	42	58	42	67
1041	**Aquifoliales**	90	122	94	130	89	120	90	123	83	103	83	123
1047*	Stem-*Aquifoliaceae*	62	81	60	99	62	80	62	81	62	76	62	81
1054	**Lamiales**	76	101	81	108	75	100	76	101	67	86	67	101
1075*	**Solanales**	75	105	81	111	75	105	77	105	66	88	66	105
1083	**Gentaniales**	65	97	70	103	64	96	65	97	57	81	57	93
1090	**Boraginales**	49	93	51	97	48	91	49	93	43	79	43	93
1093	**Garryales**	58	111	62	118	57	109	59	111	50	94	50	111
1094	**Ericales**	99	122	100	130	99	121	99	122	94	105	94	122
1098*	Ericales core	85	98	83	103	85	98	85	98	85	93	85	98
1115*	**Cornales**	85	115	84	134	85	115	85	116	86	105	86	116
1120	**Caryophyllales**	109	133	112	142	107	131	109	133	94	109	94	133
1162	**Berberidopsidales**	18	111	16	118	18	111	18	110	17	100	17	111

1163	**Santanales**	100	142	107	153	99	140	101	143	87	115	87	143
1168	**Dilleniales**	39	91	44	96	39	88	40	93	33	74	33	93
1170	**Gunnerales**	62	143	69	155	63	141	63	144	55	116	55	144
1171*	**Buxales**	99	124	96	138	99	122	99	123	99	112	99	124
1173	**Trochodendrales**	3	29	4	30	3	28	3	28	2	24	2	29
1175*	**Proteales**	108	151	110	173	108	148	107	150	108	125	107	151
1178	**Sabiales**	48	126	49	140	45	125	48	127	41	110	41	127
1179	**Ranunculales**	101	137	103	152	101	136	101	138	96	123	96	138
1182*	Stem-Menispermaceae	83	103	81	118	83	103	83	104	82	97	82	104
1193*	Monocots	144	179	159	203	141	176	141	181	123	135	123	181
1200	**Poales**	76	106	74	114	72	104	78	108	64	94	64	108
1208	**Zingiberales**	75	89	71	93	75	88	75	89	74	84	74	89
1209*	Stem-Musaceae	74	84	68	87	74	84	74	84	74	81	74	84
1210	**Commelinales**	62	99	62	106	55	98	58	100	52	90	52	100
1211*	**Arecales**	82	95	77	99	82	94	82	94	82	91	82	95
1212	**Asparagales**	94	123	100	136	93	121	93	124	85	105	85	124
1221*	Orchidaceae	18	21	13	76	18	21	18	22	18	21	18	22
1222*	**Liliales**	66	115	75	129	65	113	65	115	58	98	58	115
1224*	Discorales-Pandanales	91	131	105	152	90	128	90	131	89	108	89	131
1225	**Pandanales**	45	99	48	113	44	97	45	99	38	86	38	99
1226*	**Alismatales**	105	143	125	172	105	140	104	143	98	116	98	95
1228*	*Alisma-Potamogeton*	66	94	74	128	66	92	66	95	66	82	66	95
1232*	Araceae	76	112	88	151	76	111	76	113	76	98	76	113
1235	Magnoliidae	141	190	155	212	141	185	141	190	128	140	128	190
1237*	**Laurales**	107	134	107	153	107	132	107	133	107	121	107	134
1248*	**Magnoliales**	110	129	108	142	110	128	110	129	109	120	109	129
1256*	Stem-Canellales	128	179	143	203	127	174	128	179	117	134	117	179
1257	**Piperales**	106	152	115	171	105	148	106	152	95	120	95	152
1260*	Stem-*Saururus*	44	63	43	84	44	63	44	63	44	59	44	63
1264	**Canellales**	67	135	70	154	67	131	67	134	61	112	61	135
1268*	**Chloranthales**	92	113	88	126	92	113	92	113	92	107	92	113
1271	**Austrobaileyales**	119	167	119	180	119	165	119	167	115	135	115	167
1272*	Stem-Schisandraceae	107	132	105	148	107	132	107	132	107	121	107	132
1273*	Stem-*Illicium*	85	103	80	112	85	104	85	103	85	101	85	104
1275	**Nymphaeales**	128	189	129	198	128	185	129	188	119	140	119	189
1276*	Stem-*Cabombaceae*	110	128	107	138	110	128	110	128	109	119	109	128
1278*	Nymphaceae	93	116	91	126	93	116	93	116	94	111	93	116
1279*	Acrogymnosperms	308	357	307	342	308	355	307	351	307	345	307	357
1280*	Conifers	274	327	254	318	271	325	264	322	265	321	265	327
1282*	Gnetales	116	160	113	184	115	161	116	166	115	151	115	166
1284*	*Ginkgo-Cycas*	264	317	263	310	264	318	263	313	263	311	263	317
1286*	Monilophytes	383	413	382	411	383	409	383	415	383	415	383	413
1287*	Lycophytes	390	430	395	452	391	432	389	443	390	444	389	443

Note: Nodes are numbered as in Figure B.2. The 52 calibrated nodes are represented by (*) and nodes in bold characters represent major angiosperm orders. Posterior times are the 95% HPD CI, estimated using the HKY85+Γ_5 substitution model, IR rate model, with the 83 genes subdivided into three partitions: (1) 1st and 2nd codon positions for plastid genes; (2) 1st and 2nd codon positions for mitochondrial genes; and (3) nuclear RNA genes. Composite: 95% high posterior density credibility interval (HPD CI) is a composite of the 95% HPD credibility intervals across all calibration strategies, except calibration strategy B (SB) (Figure 3.4).

3.4.1 Impact of partition strategy on divergence time estimates

Divergence time estimation can also be affected by the manner in which the molecular sequence alignment is partitioned (Zhu et al., 2015). Thus, we considered three different partition schemes. In the first (3P), the sequence alignment was subdivided into three partitions (excluding 3rd codon positions): (i) 1st and 2nd codon positions for plastid genes; (ii) 1st and 2nd codon positions for mitochondrial genes; and (ii) nuclear RNA genes. In the second (1P) these partitions were concatenated and

analyzed as a single partition. Our third partition strategy (MP) was a mixed alignment divided into plastid proteins, mitochondrial proteins and nuclear RNA genes. Divergence time analysis using partition scheme 1P yielded the least precise estimates (Table B.4) and the posterior mean age estimates are the least compatible with the other partition schemes (Figure 3.4a and Table B.4). This is because the posterior precision of time estimates is strongly influenced by the number of partitions (Zhu et al., 2015). Estimates using 3P and MP are more precise and much more consistent with one another (Figure 3.4b and Table B.4), though the improvement is more marked between one partition and three partitions, than between three nucleotide partitions and three hybrid partitions, suggesting that 3P achieves the best trade-off between increasing analytical complexity and accuracy.

3.4.2 Impact of rate model on divergence time estimates

The rate models can also affect divergence time estimation when the molecular clock is seriously violated (dos Reis et al., 2015), as it is among angiosperms (Beaulieu et al., 2015). When the clock is violated, rates calculated in one part of the phylogeny serve as a poor proxy for estimating divergence times in other clades. To assess the effect of this uncertainty, we estimated divergence times for tracheophytes assuming an autocorrelated (AR) rates model under calibration strategy A. In attempting to encompass the uncertainty in the rate drift model we consider here the spread of node age estimates that arise from both rate models (Figure 3.4c). Our results show that the rate model has a large effect, with the autocorrelated-rates model producing older estimates for shallow nodes and younger estimates for deep nodes, in comparison to the independent rates model, where a few nodes, especially the deep nodes, are younger (Figure 3.4c and Table B.4). Moreover, we tested a series of informative priors on the overall rate based on the rough rate estimates mentioned above. However, these priors did not affect time estimates noticeably, possibly because a large number of fossil calibrations constrain the time prior.

3.4.3 Bayes factor calculation for clock model selection

To evaluate the performance of different molecular clock models we used Bayes factors to calculate the ratio of the marginal likelihoods for two models under comparison. The marginal likelihood is complicated to estimate, however, recently

methods such as path-sampling (thermodynamic integration) and stepping-stones have been developed for phylogenetics (Lartillot & Philippe 2006; Lepage et al., 2007; Linder et al., 2011; Xie et al., 2011; Rannala & Yang 2017). Here, we implement the path-sampling approach to calculate the marginal likelihoods for the strict clock (SC), IR and AR models. Because the thermodynamic integration is computationally expensive (Rannala & Yang 2017), we calculated the marginal likelihood for the SC, IR and AR models using a smaller dataset of ten tracheophyte species (*Huperzia*, *Psilotum*, *Ginkgo*, *Amborella*, *Nymphaea*, *Acorus*, *Calycanthus*, *Platanus*, *Oxalis* and *Cornus*), for the 4 partitions analyzed (Table S2). The results are shown in Table 1. The IR model always has the highest marginal likelihood, with the posterior model probability > 90% in all datasets. Therefore, we conclude that overall, the IR model is the most appropriate model of rate variation on the tracheophyte data analyzed here, and the divergence times calculated under the IR model should be preferred. We expect that these results should apply to the larger datasets used in the estimation of divergence times.

Table 3.3: Bayesian model selection of rate model.

Dataset	Clock model	Log Marginal *L*	BF	*P*
Plastid 1st and 2nd c.p.	SC	–141,585.67	5.1×10^{-274}	5.05×10^{-274}
	IR	**–140,956.40**	**—**	**0.991**
	AR	–140,961.16	0.009	0.009
Mitochondrial 1st and 2nd c.p.	SC	–13,776.34	7.86×10^{-29}	7.79×10^{-29}
	IR	**–13,711.64**	**—**	**0.991**
	AR	–13,716.36	0.009	0.009
Nuclear RNA	SC	–17,534.24	2.15×10^{-41}	2.03×10^{-41}
	IR	**–17,440.60**	**—**	**0.944**
	AR	–17,443.43	0.059	0.056
Concatenation (pl1&2, mt1&2, nucRNA)	SC	–173,121.00	1.03×10^{-297}	1.02×10^{-297}
	IR	**–172,437.16**	**—**	**0.988**
	AR	–172,441.60	0.012	0.012

SC: Strict clock; IR: Independent rates; AR: Auto-correlated rates. The age of the root is fixed to one (i.e. we use a 'B(0.99, 1.01)' calibration on the root in MCMCTree). The rate prior used is G(2, 10). The prior on σ^2 is G(2, 4) in all cases. The model with the highest posterior probability in each dataset is shown in bold type.

3.4.4 Impact of diversification model on divergence time estimates

We also explored the impact of the birth-death process used to specify the prior of times on divergence time estimation. The parameters of the birth-death process with

species sampling were fixed at λ =1, μ = 1, ρ = 0, which generates uniform node ages. We assessed uncertainty by adjusting parameters λ, μ and ρ such that the kernel has an L shape (λ = 1, μ = 4, ρ = 0.1), giving a tree with long internal branches (BD1), or an inverted L shape (λ = 4, μ = 1, ρ = 0.0001), giving a tree with long terminal branches (BD2). The results of these two parameter sets are almost identical to those from the original setting (Figures 3.4d and 3.4e), suggesting that parameter selection for the birth-death does not have a significant impact on divergence time estimates for this dataset.

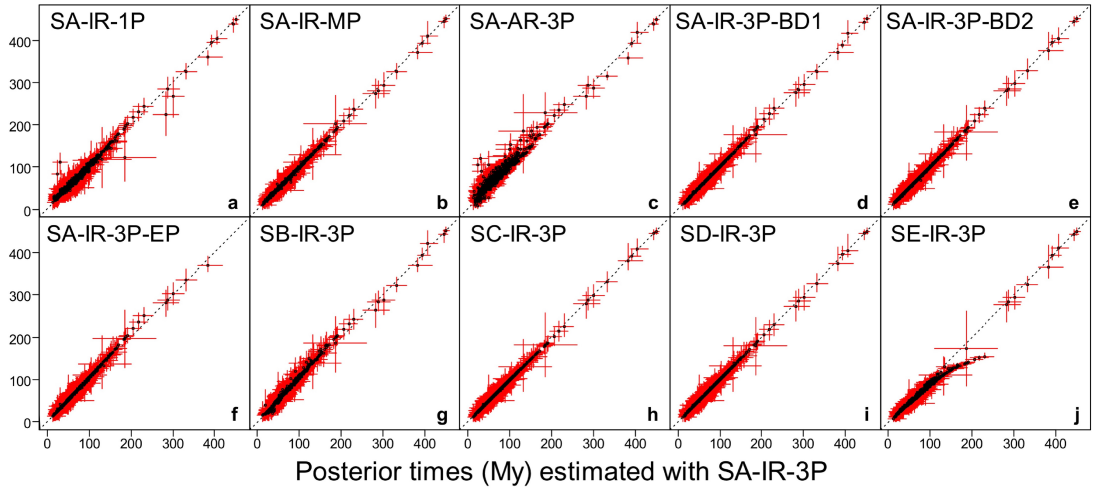


Figure 3.4: Sensitivity of times estimates to the number of partitions, rate model, birth-death process, exclusion of lycophytes + monlilophytes, and fossil calibrations. The posterior mean times (black dots) and 95% CIs (red lines) of 643 nodes under calibration strategy A (SA), independent rates (IR) model, and gene alignments and 3 partitions are plotted against **(a)** estimates using 1 partition, **(b)** mixed partitions, **(c)** autocorrelated rates (AR) model, **(d)** birth-death parameters adjusted to generate a tree with long internal branches and short tip branches (BD1) and **(e)** large node ages with nodes close to the root (BD2), **(f)** excluding ferns and lycophytes (EP), **(g)** calibration strategy B (SB), **(h)** calibration strategy C (SC), **(i)** calibration strategy D (SD) and **(j)** calibration strategy E (SE).

3.4.5 Impact of outgroup sampling on divergence time estimates

Finally, we considered the impact of the choice of outgroups on divergence time estimation. We included several outgroups to seed plants so that we could consider the timing of angiosperm origin in the context of land plant diversification as a whole. However, the ferns and lycophytes are distantly related clades comprised of long branches, and may therefore have biased our estimates. We explored the effect of including distantly related outgroups (tracheopytes dataset) and of excluding lycophytes and ferns (EP dataset). The results (Fig. 4f and Table S4) show that

including lycophytes and ferns did not have a strong effect on the posterior time estimates, although their exclusion did result in increased ages for some intermediate clades.

3.5 Discussion

Overall, the estimated divergence times for angiosperm clades are robust to variation in models and parameters within our model of land plant phylogeny. These include the birth-death prior and the prior for rate parameters under the rate drift model. The main factors affecting the estimates in this study are data partitioning, fossil calibration uncertainty and the discrepancy between the user specified time prior and the effective time prior, and the rate-drift model. Regarding relaxed clock models, an applicable result from our study is the finding that the IR model fits the tracheophyte data better than the AR relaxed-clock model. In agreement with our results, Linder, et al. (2011) found, using Bayes factors, that the IR model fitted angiosperm and eudicot data better than the AR model. In the IR model the variance of the rate is independent to the time, thus the rate can undergo large shifts -- depending on the value of σ^2 -- including for adjacent branches. On the contrary, in the AR model the variance depends on the time and hence the model penalizes large rate variation over short time intervals but allows rate to vary nearly freely among distant clades. However, the variance increases linearly with the time and in analyses of deep phylogenies this might lead to excessively high rate shifts. Therefore, the AR model might be more suitable for the analysis of closely-related species while the IR model for divergent species and large phylogenies. However, further research is still needed to understand which clock model is the most biologically realistic and appropriate for real data analysis (Lepage et al., 2007; Ho & Phillips 2009; Linder et al., 2011).

None of our component analyses provides an accurate timescale for angiosperm evolution since each one controls for a different source of uncertainty. Rather, it is necessary to integrate these uncertainties into a single timescale (Fig. 3.5 and Table 3.2). This allows us to conclude that the crown Tracheophyta and crown Euphyllophyta originated in the late Ordovician – early Silurian interval (458-442 Ma and 455-427 Ma, respectively) and Spermatophytes within the latest Silurian – early Carboniferous (422-340 Ma. Crown-angiosperms originated within the late Permian – latest Jurassic interval (256-149 Ma) whereas the crown clades of magnoliids,

monocots and eudicots diverged between the early Jurassic and early Cretaceous (190-128, Ma, 181-123 Ma and 188-129 Ma, respectively), and the two main lineages of eudicots, the asterids and rosids, originated between the latest Jurassic and middle Cretaceous (146-107 Ma and 160-117 Ma, respectively). Whereas the age estimates for non-angiosperms were close to age estimates based on a literal reading of the fossil record, the conflicts between the molecular estimates and the fossil record were greater within angiosperms.

Recent studies provide a huge spread of molecular clock estimates for the origin of crown angiosperms (Bell et al., 2005; Magallón & Castillo 2009; Bell et al., 2010; Smith et al., 2010; Clarke et al., 2011; Magallón et al., 2013; Magallón 2014; Zanne et al., 2014; Zeng et al., 2014; Beaulieu et al., 2015; Magallón et al., 2015; Foster et al., 2016; Murat et al., 2017), ranging from the Permian (Bell et al., 2010; Smith et al., 2010; Beaulieu et al., 2015) to the Lower Cretaceous (Bell et al., 2005; Magallón 2009; Bell et al., 2010; Magallón et al., 2015), covering on range 122-270 Ma. Our integrated timescale, which encompasses all of the unconstrainable sources of uncertainty we addressed (Fig. 3.5 and Table 3.2), estimates crown angiosperms to have diverged in the interval 149-256 Ma, fully within the range of previous estimates. Apart from a range of methodological differences, it is the manner in which the palaeontological data are interpreted to implement fossil constraints, that accounts for many differences between our estimates and those obtained in previous studies. For example, analyses that yield Cretaceous estimates for the origin of angiosperms have used a Cretaceous point calibration or a concentrated calibration density, under the assumption that the age of crown-angiosperms is known almost without error. (Magallón 2009; Magallón et al., 2015). In general, recent molecular clock studies obtained estimates suggesting a Triassic origin of angiosperms. Hence these molecular estimates raise the possibility that the oldest crown angiosperm fossils are still undiscovered, or at least unidentified.

The integrative timescale that we have derived from our analyses is less precise than many previous analyses, however, because it integrates most sources of unconstrainable uncertainties, it can be considered the most accurate obtained until now. These uncertainties inevitably result in estimates that encompass clade ages in closer accord with fossil evidence, but for key clades like the angiosperms

themselves, it still does not encompass clade ages proposed based on analyses of the fossil record alone. Thus, though we specifically considered variables that might yield younger age estimates, like increased ingroup and outgroup taxon sampling, and probabilistic time priors skewed towards clade minimum age calibrations, the results of our analyses are consistent with previous analyses in implying a cryptic interval of early angiosperm evolution. There appears little scope for credibly finding closer agreement between molecular clock estimates and the fossil record of angiosperm origin.

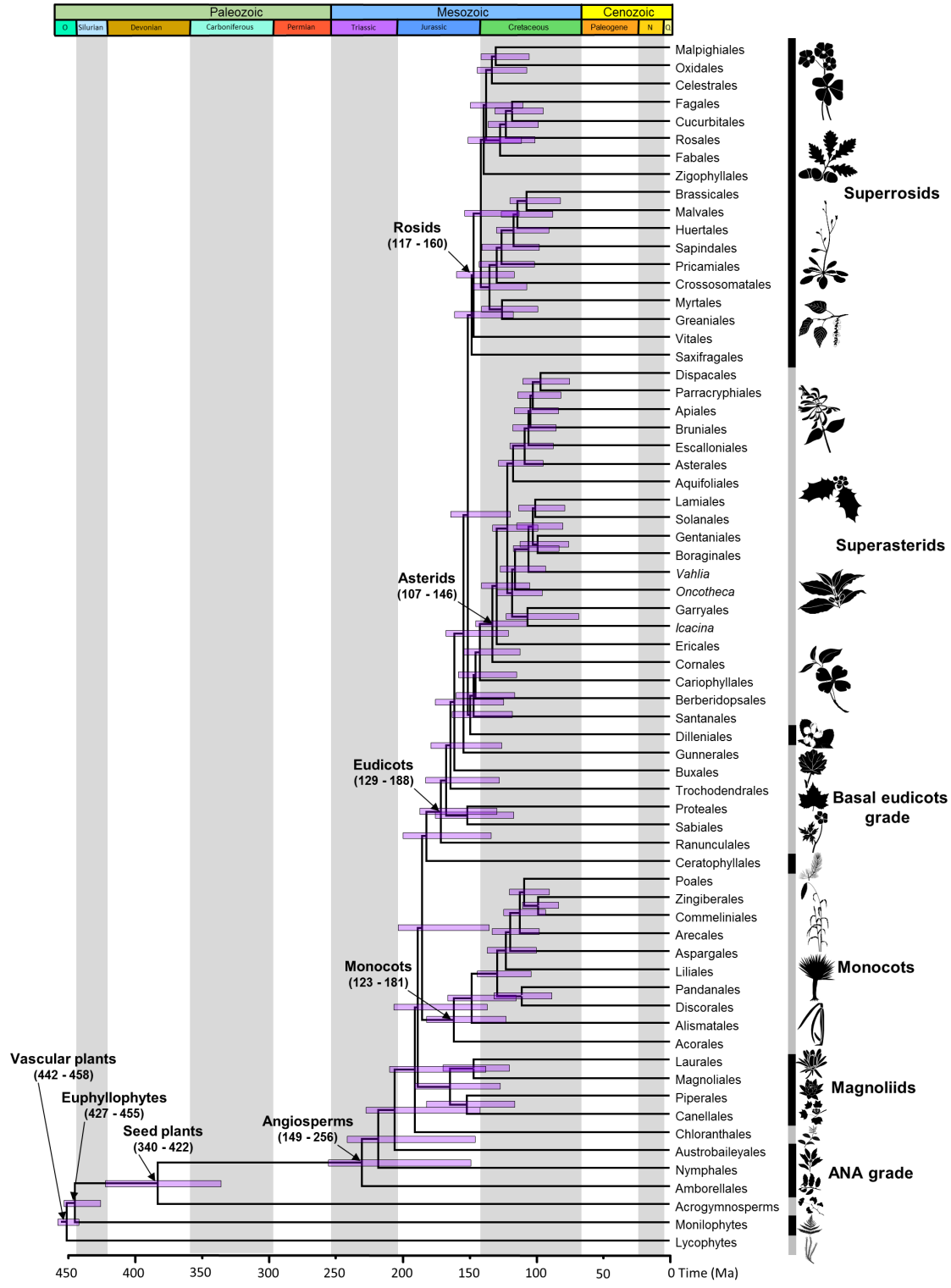


Figure 3.5: The time tree of tracheophytes encompassing uncertainty of calibration strategies Summary chronogram for tracheophytes with terminals collapsed to represent angiosperm orders. Node ages are plotted at the posterior mean for calibration strategy A (SA), 3 partitions (3P), independent rates model (IR) and HKY85+Γ5 substitution model. The Node bars are composites extending from the minimum 2.5% HPD limit to the maximum 97.5% limit across all calibration strategy analyses (excluding results from calibration strategy B).

The discordance between molecular clock estimates and unequivocal fossil evidence of crown angiosperms implies a cryptic interval to their early evolutionary history spanning at least 23 Myr, and as much as 121 Myr, in which angiosperms existed but are unrepresented in the fossil record. However, the apparent mismatch may be more perceived than real. Though the early fossil record of angiosperms has been interpreted to reflect an orderly and incrementally phased environmental invasion (Hickey 1997; Friis et al., 2010; Coiffard et al., 2012; Doyle 2012), this pattern may be an artefact imposed by the non-uniformity of the rock record on the fossil record of all terrestrial clades (Benson et al., 2013). Furthermore, while the earliest unequivocal evidence of angiosperms, based on (Fischer's rule) tricolpate pollen, can be constrained minimally to the Barremian, this actually evidences the establishment of the eudicot lineage, which is remote from the angiosperm crown ancestor (Doyle & Hotton 1991; Clarke et al., 2011). Monosulcate pollen, like that produced by early-branching lineages of extant angiosperms, is known at least as far back as the Valanginian (Brenner 1996), and pollen exhibiting subsets of definitive crown-angiosperm characters are known as far back as the Middle Triassic (Crane 1987; Cornet 1989; Doyle & Hotton 1991; Taylor & Taylor 2009; Hochuli & Feist-Burkhardt 2013), but these are difficult to discriminate from pollen produced by stem-angiosperms or gymnosperms (Doyle & Hotton 1991) and, hence, they have not been used to constrain divergence time analyses. There are also claims of pre-Cretaceous crown-angiosperms based on macrofossil evidence. While the age of the angiosperm macrofossil genus *Archaefructus* (Sun et al., 2002; Friis et al., 2003) has been revised from Jurassic to Cretaceous (Chang et al., 2009). Other putative pre-Cretaceous angiosperm fossils are more securely dated but their interpretation requires further attention (Crane et al., 1995; Taylor & Taylor 2009; Friis et al., 2011; Doyle 2012; Liu & Wang 2016a; Liu & Wang 2016b; Herendeen et al., 2017). However, demonstration that pre-Cretaceous seed plant macrofossils fail to exhibit conclusive evidence of crown-angiosperm affinity (e.g Herendeen et al., 2017) is not the same as demonstrating that they are not crown-angiosperms, or that crown angiosperms diverged in the Cretaceous. This false logic is invariably based in absence of evidence of 'key characters' rather than evidence of their absence (e.g Herendeen et al., 2017), and couched within the increasingly dated parsimony-based phylogenetic framework (Wright & Hillis 2014; O'Reilly et al., 2016; Puttick et al., 2017) used to inferring both seed plant relationships and also the phylogenetic distribution of characters.

Symptomatically, much of the controversy over seed plant relationships is rooted in the false-precision of parsimony-based phylogenetic analyses of morphological characters (O'Reilly et al., 2016; Puttick et al., 2017). At the least, the hypotheses of character evolution used to discriminate stem and crown angiosperm fossil taxa should be reviewed within a probabilistic framework that can better accommodate the uncertainty associated with such inference. However, it may be more appropriate to reconsider the phylogenetic position of critical fossil taxa using likelihood models of character evolution to accommodate phylogenetic uncertainty (Wright & Hillis 2014; O'Reilly et al., 2016; Puttick et al., 2017) since discriminating between a stem- and crown-angiosperm affinity of all pre-Cretaceous claims may be the only way in which molecular estimates for the origin of flowering plants are going to achieve accuracy and precision. Contemporary interpretations of the palaeobotanical record do not present any material constraints on extent of the pre-Cretaceous history of crown angiosperms.

Thus, the results of our analyses allow us to reject the hypothesis that crown-angiosperms originated in the Cretaceous and, as such, allow us to reject the extreme hypothesis of KTR, or an explosive diversification of flowering plants fully within the Cretaceous (Cascales-Minana et al., 2016). However, our results remain compatible with a more general hypothesis of a KTR in that diversification of the major groups of angiosperms occurred later (150-100 Ma), contemporaneous with the explosive diversification of derived lineages of insects (Misof et al., 2014), seed-free land plants (Schneider et al., 2004; Feldberg et al., 2014; Laenen et al., 2014), and within the interval in which the fossil record reflects flowering plants to have risen to ecological dominance in terrestrial communities.

3.6 Conclusions

The timescale of angiosperm origin and diversification is one of the iconic clock-rock controversies. Furthermore, the timescale of angiosperm diversification varies broadly among different molecular analyses. This is expected given that transforming molecular distances into geological divergence times is challenging. There are many methodological variables in previous molecular analyses, that are known to affect the accuracy and precision of divergence time estimates. We aimed to control some limitations from previous studies (e.g. low taxon sampling, limited sequence data,

insufficient outgroup lineages or a combination of these shortcomings), while also controlling for several sources of uncertainty. However, our attempts to control for them have resulted in greater uncertainty in the molecular timescale, itself allowing for closer agreement, especially for the diversification of clades within crown angiosperms. The timing of crown angiosperm diversification remains offset from traditional interpretations of the fossil evidence, by as little as a couple of tens of millions of years or as much as a hundred million years or more. The discovery of equivocal microfossil and macrofossil angiosperm remains in the Jurassic and even Triassic, together with evidence of biases in the rock record, suggest that this mismatch may be real. However, it also suggests that agreement between clock and rock estimates may be found in quantification of the artefact imposed by the rock record on the fossil record, and the further analysis of equivocal pre-Cretaceous remains. Nevertheless, we also recognize that a closer match between molecular estimates and the fossil record may be eventually achieved, as dating methods are improved to deal with high rate heterogeneity. Even though the extreme interpretation of a KTR appears incorrect, the diversification of speciose clades among crown angiosperms does appear to coincide with that of herbivores and pollinators and their predators, corroborating a more general hypothesis of a KTR. This underlines the power of the complementary nature of molecular and palaeontological data and approaches to inferring evolutionary timescales and establishing a deeper understanding of clade dynamics in deep time.

4 Comparison of different strategies for using fossil calibrations to generate the time prior in Bayesian molecular clock dating

4.1 Abstract

Fossil calibrations are the utmost source of information for resolving the distances between molecular sequences into estimates of absolute times and absolute rates in molecular clock dating analysis. The quality of calibrations is thus expected to have a major impact on divergence time estimates even if a huge amount of molecular data is available. In Bayesian molecular clock dating, fossil calibration information is incorporated in the analysis through the prior on divergence times (the time prior). Here, we evaluate three strategies for converting fossil calibrations (in the form of minimum- and maximum-age bounds) into the prior on times, which differ according to whether they borrow information from the maximum age of ancestral nodes and minimum age of descendent nodes to form constraints for any given node on the phylogeny. We study a simple example that is analytically tractable, and analyze two real datasets (one of 10 primate species and another of 48 seed plant species) using three Bayesian dating programs: MCMCTree, MrBayes and BEAST2. We examine how different calibration strategies, the birth-death process, and automatic truncation (to enforce the constraint that ancestral nodes are older than descendent nodes) interact to determine the time prior. In general, truncation has a great impact on calibrations so that the effective priors on the calibration node ages after the truncation can be very different from the user-specified calibration densities. The different strategies for generating the effective prior also had considerable impact, leading to very different marginal effective priors. Arbitrary parameters used to implement minimum-bound calibrations were found to have a strong impact upon the prior and posterior of the divergence times. Our results highlight the importance of inspecting the joint time prior used by the dating program before any Bayesian dating analysis.

Keywords: Bayesian inference, molecular clock dating, divergence times, fossil calibration, time prior.

4.2 Introduction

Bayesian inference has become the methodology of choice for molecular clock dating of species divergences because it provides a natural framework for incorporating different sources of information (e.g., from fossils and molecules) (dos Reis et al., 2016). In a Bayesian dating analysis, one would ideally summarize the relevant prior evidence about species divergence times (say, from the fossil record, geological events, etc.) in a multidimensional joint prior of ages for all nodes on the phylogeny (called the time prior). However, specifying high-dimensional priors with complex correlation structures is a notoriously difficult task, and furthermore, our knowledge of the fossil evidence and of how it informs the species divergence times is very imprecise. The current practice is for the paleontologist to specify minimum- and maximum-age constraints on certain nodes on the tree based on the fossil evidence (Thorne et al., 1998; Kishino et al., 2001; Benton et al., 2009; Ho & Phillips 2009). Such *user-specified* fossil calibrations are then used by the Bayesian dating program to construct the time prior, with the distribution of the ages of non-calibration nodes supplanted by a branching-process model (e.g., a birth-death process) (Yang & Rannala 2006). The user-specified calibration densities are assigned to single nodes on the tree and often do not satisfy the requirement that any ancestral node should be older than its descendants, and thus the dating software must 'truncate' the calibration densities to satisfy this constraint. We refer to the resulting prior of node ages used by the dating software as the *effective prior*, and this may be very different from the original user-specified calibration densities (Inoue et al., 2010; Warnock et al., 2015). Furthermore, Bayesian dating programs such as MultiDivTime (Thorne et al., 1998), MCMCTree (Yang 2007), BEAST2 (Bouckaert et al., 2014) and MrBayes (Ronquist et al., 2012b) use different procedures to combine calibration densities with the birth-death process model to generate the time prior, so that different programs may produce very different time priors from the same user-specified fossil calibrations (Inoue et al., 2010).

Thus, users of dating software are encouraged to run the Markov Chain Monte Carlo (MCMC) algorithm without molecular data to generate the time prior used by the program and to inspect it to ensure that it is a reasonable representation of the fossil evidence. A cross-validation method for assessing the quality of calibrations, based on the consistency between fossils and between fossils and molecules, has also been

proposed (Near et al., 2005). This was noted to sometimes lead to the selection of calibrations of poor reliability (Marshall 2008; Benton et al., 2009; Warnock et al., 2015). The problem appears to be partly due to the fact that fossil-calibration constraints provided by the paleontologist are "over-interpreted" by the Bayesian dating program. For example when fossil evidence suggests that the age of a clade is between 50Ma and 100Ma, current dating software incorporates that information by assigning a uniform distribution, $t \sim U(50, 100)$, implying, for example, $P\{50 < t < 60\} = P\{90 < t < 100\}$. Such probabilistic statements about the true age may not be intended by the paleontologist. However minimum and maximum bounds alone, in the form of $50 < t < 100$, are insufficient to permit a Bayesian dating analysis: a full statistical distribution for the true age has to be specified.

The way that the fossil-based bounds on node ages are converted into statistical distributions in a dating analysis may thus have an important impact on the posterior time estimates. Consider the unbalanced 5-species phylogeny of Figure 4.1. Suppose that fossil evidence suggests that the age of node 4 should be at least 10 Myrs, while the age of the root is at most 100 Ma, with $t_4 > 10$ and $t_1 < 100$ (Figure 4.1). Three simple strategies appear possible to construct the calibration densities. In strategy 1 (st1), we apply a minimum-bound calibration on t_4, by using a decay function from 10 Ma to ∞ (such as the offset-exponential), while the age of the root may be assigned a uniform distribution $t_1 \sim U(0, 100)$. Ages of the non-calibration nodes (t_2 and t_3) have densities specified by the birth-death process. In strategy 2 (st2), we propagate the minimum and maximum bounds to all calibration nodes: the root acquires the minimum bound from node 4, while node 4 inherits the maximum age of the root, so that both nodes have joint bounds: $t_4 \sim U(10, 100)$, and $t_1 \sim U(10, 100)$. In strategy 3 (st3), we propagate the minimum and maximum bounds to all nodes on the phylogeny, so that $t_i \sim U(10, 100)$ for $i = 1, 2, 3$ and 4. In all three strategies, the dating program will automatically apply a truncation so that $t_4 < t_3 < t_2 < t_1$. Different programs use different procedures to perform the truncation and to combine the calibration densities with the branching process model (Inoue et al., 2010). As a result, the three strategies should lead to different time priors, and the different programs will also differ even for the same strategy. For simple cases, it is possible to calculate analytically the resulting marginal priors for the node ages after truncation. However, for large phylogenies with dozens of fossil calibrations, analytical calculation is impossible, and

the user needs to estimate the prior by running the Bayesian MCMC program without sequence data.

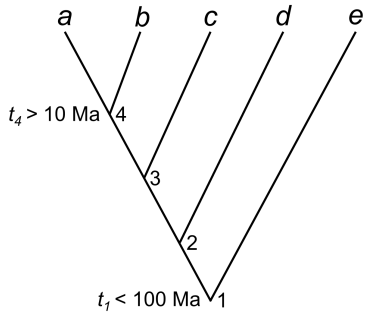


Figure 4.1: A five species phylogeny used in the analytical example of fossil calibration strategies.

Here we study how the different calibration strategies affect the time prior and the posterior time estimates. We examine two approaches used by Bayesian dating programs to combine calibration densities with the branching process to form a prior density for all node ages (the time prior): the *conditional construction* used by MCMCTree (Yang & Rannala 2006) and the *multiplicative construction* used by BEAST (Bouckaert et al., 2014) and MrBayes (Ronquist et al., 2012b) (see Heled and Drummond, 2015). We study a simple example that is analytically tractable, and then analyze two real datasets: one of 10 primate species, and another of 48 seed plant species. We show that the different calibration strategies as well as truncation have significant impacts on the time prior and the resulting posterior time estimates. We discuss the implications of our results and give recommendations for the construction of reasonable time priors.

4.3 Material and methods

4.3.1 Fossil calibrations and the time prior

We consider three types of constraints on a node age based on the fossil evidence: minimum-age bound, maximum-age bound, and joint (maximum- and minimum-age) bounds (Figure 4.2). These are implemented in different Bayesian dating programs using different approaches.

Minimum-age calibrations (Figure 4.2a). In MCMCTree, a minimum bound is represented using a truncated Cauchy distribution, denoted L(t_L, p, c, p_L) (Inoue et al., 2010) Here t_L is the minimum age bound, p determines how far the mode of the distribution is from the minimum, c determines how sharply the distribution decays to zero, and p_L is the left tail probability (i.e. the probability that the minimum bound is violated). Smaller values of p and c give a more concentrated calibration density, with the true age being closer to the minimum age. For example, p = 0.1 means the mode of the distribution is at $(1 + p)t_L = 1.1t_L$. Here we used p = 0.1, c = 0.1, and p_L = 0.01.

In MrBayes and BEAST2, minimum bounds are represented using an offset-exponential distribution (Ronquist et al., 2012b; Bouckaert et al., 2014; Heled & Drummond 2015). If y has an exponential distribution with rate parameter θ or mean $1/\theta$, then $t = y + t_L$ has an offset-exponential distribution with parameters θ and t_L, with mean $\theta^{-1} + t_L$. A large θ means that the true age is likely to be close to t_L. In this study, we used $\theta = 10/t_L$, so that the mean of the distribution is 1.1 t_L.

Maximum-age calibrations (Figure 4.2b). Maximum bounds are represented by a uniform distribution U ~ (0, t_U), where t_U is the maximum age. Bounds are hard in BEAST2 and MrBayes and are soft in MCMCTree, with p_U to be the error probability that the bound is violated.

Joint (minimum- and maximum-age) calibrations (Figure 4.2c). Joint bounds are represented by a uniform distribution U(t_L, t_U) in all three programs. Again, bounds are hard in BEAST2 and MrBayes, and soft in MCMCTree, which assigns p_L and p_U as the error probabilities for violations of the bounds (Yang & Rannala 2006). We use p_L = 0.01 and p_U = 0.05.

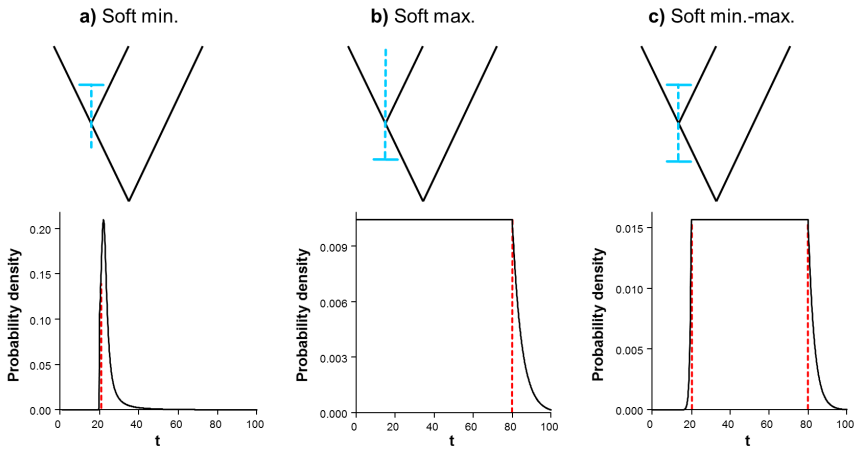


Figure 4.2: Probability densities for describing uncertainties in fossil calibrations. Probability densities for describing uncertainties in fossil calibrations: **(a)** soft minimum bound represented by a shifted-exponential distribution specified as t_L = 20, p = 0.1, c = 0.1, p_L = 0.01; **(b)** soft maximum bound specified as "t_U = 80, p_R = 0.05"; and **(c)** soft lower and upper bound, specified as "t_L = 20, t_U = 80, p_L = 0.01, p_U = 0.05".

4.3.2 Calibration strategies to generate the time prior

The calibration strategies are different ways of generating the effective prior given the fossil bounds on the calibration nodes on the phylogeny. We consider three strategies.

Calibration strategy st1: Minimum and maximum constraints were applied to calibration nodes as given, without propagating onto other nodes.

Calibration strategy st2: Minimum and maximum constraints are propagated onto all calibration nodes, so that every calibration node has joint minimum and maximum bounds, represented by a uniform distribution. In other words, if a calibration node lacks a minimum bound, the minimum bound of its oldest descendent node is used, and if a calibration node lacks a maximum bound, the maximum bound of its youngest ancestor is used.

Calibration strategy st3. This is like st2 but minimum and maximum bounds are propagated onto all interior nodes on the phylogeny, so that every node has a pair of joint bounds.

The rooted tree topology was fixed in all analyses. This is a requirement for MCMCTree and we did the same for BEAST2 and MrBayes to avoid the confounding

effects of alternative phylogenies. A constraint on the root is required in MCMCTree (Yang & Rannala 2006) and MrBayes (Ronquist, et al. 2012b). BEAST2 does not require a constraint on the root, one or more calibrations on internal nodes may be sufficient (Heled & Drummond 2012, 2015). The Bayesian analysis requires a prior on the ages of all nodes on the tree. The birth-death branching process is used to provide the prior distribution for the non-calibration nodes, which is combined with the effective prior for the calibration nodes after the truncation, to generate the time prior. Two procedures have been used to achieve this in the current dating programs.

The Bayesian analysis requires a prior on the ages of all nodes on the tree. The birth-death branching process is used to provide the prior distribution for the non-calibration nodes, which is combined with the effective prior for the calibration nodes after the truncation, to generate the time prior. Two procedures have been used to achieve this in the current dating programs.

In MCMCTree, the so-called *conditional construction* is used (Yang & Rannala 2006). Let t_C be the ages of the calibration nodes, and $t_{\bar{C}}$ be the ages of the non-calibration nodes. In the example of Figure 4.1, $t_C = \{t_1, t_4\}$ while $t_{\bar{C}} = \{t_2, t_3\}$. The conditional construction gives the density of all node ages as

$$f(t_C, t_{\bar{C}}) = f(t_C) \cdot f_{\mathrm{BD}}(t_{\bar{C}} \mid t_C), \tag{4.1}$$

where $f(t_C)$ is the effective prior on the ages of the calibration nodes, given by the user-specified calibration densities after truncation, while $f_{\mathrm{BD}}(t_{\bar{C}} \mid t_C)$ is the conditional density of the non-calibration nodes given the calibration node ages, specified by the birth-death-sampling process (Yang & Rannala 1997).

Both BEAST2 and MrBayes use the so-called *multiplicative construction*, in which the birth-death process density for all node ages is multiplied with the densities for the calibration nodes to generate the time prior (Heled & Drummond 2012, 2015).

$$f(t_C, t_{\bar{C}}) \propto f(t_C) \cdot f_{\mathrm{BD}}(t_C, t_{\bar{C}}) = f(t_C) \cdot \left[f_{\mathrm{BD}}(t_C) \cdot f_{\mathrm{BD}}(t_{\bar{C}} \mid t_C) \right] \quad (4.2)$$

As the density of t_C occurs twice in eq. 4.2, this construction "does not follow the rules of probability calculus" (Heled and Drummond, 2012). Here we treat both constructions as heuristic methods for converting user-specified constraints into the time prior.

4.3.3 Analysis of a simple example with five species

We use a simple and analytically tractable case of five species (Figure 4.1) to explore the different approaches to constructing the time prior (the prior for all node ages). Nodes 1 and 4 are calibration nodes, with the fossil constraints $t_1 < 100$ Myrs and $t_4 > 10$, while t_2 and t_3 are non-calibration nodes, for which the densities are provided by a branching process such as the birth-death-sampling process. As the birth-death process has no beginning and no ending, it is necessary to condition the process either on the time of origin, or the age of the root, or on the number of sampled extant species (Yang & Rannala 1997). Here we condition on both the number of sampled extant species and the age of the root, as in Yang and Rannala (1997). We fix the parameters in the model at $\lambda = \mu = 1$ and $\rho = 0$, so that the ages of the nonroot nodes are order statistics from a uniform kernel (Yang & Rannala 1997). In other words, given the root age t_1, node ages t_2, t_3 and t_4 can be generated by sampling three independent random variables from U(0, t_1) and then ordering them. The joint distribution is

$$f_{BDS}(t_4 t_3 t_2 | t_1) = \frac{3!}{t_1^3}, \quad 0 < t_4 < t_3 < t_2 < t_1, \quad (4.3)$$

this is equivalent to the Dirichlet time prior used by Thorne et al. (1998).

Calibration strategy 1 (st1). We consider the *conditional* construction used by MCMCTree first (Yang & Rannala 2006). The calibration density for t_1 (the root age) is the uniform distribution

$$f_C(t_1) = \frac{1}{t_1}, \quad 0 < t_1 < t_U \quad (4.4)$$

with t_U = 100, while that for t_4 is the offset-exponential

$$f_C(t_4) = \theta\, e^{-\theta(t_4 - t_L)}, \quad t_L < t_4 < \infty, \tag{4.5}$$

where t_L = 10 and we choose $\theta = 1/t_L$ so that the mean is $2t_L$ = 20 Ma.

Multiplying those user-specified calibration densities and removing the unfeasible region (where $t_4 > t_1$) by truncation leads to the effective prior used by the dating program

$$f_c(t_4, t_4) = \frac{1}{k_1}\theta\, e^{-\theta(t_4 - t_L)}, \quad t_L < t_4 < t_1 < t_U, \tag{4.6}$$

where $k_1 = \int_L^{t_U}\int_{t_L}^{t_1} \theta\, \mathrm{e}^{-\theta(t_4 - t_L)} \cdot \frac{1}{t_U}\, \mathrm{d}t_4 \mathrm{d}t_1$ = 0.80001 is a normalizing constant, to ensure that $f_C(t_1, t_4)$ integrates to 1.

Under the birth-death-sampling process model, with $\lambda = \mu = 1$ and $\rho = 0$, the joint density for t_2 and t_3, conditioned on the calibration node ages (t_1 and t_4), is given by the fact that t_2 and t_3 are order statistics from U(t_4, t_1), with density

$$f_{BDS}(t_2, t_3 | t_1, t_4) = 2/(t_1 - t_4)^2, \quad t_4 < t_3 < t_2 < t_1. \tag{4.7}$$

The effective time prior or the joint density for all node ages is thus

$$f(t_1, t_2, t_3, t_4) = f_c(t_1, t_4) f_{BDS}(t_2, t_3 | t_1, t_4) = \frac{1}{k_1}\theta\, \mathrm{e}^{-\theta(t_4 - t_L)} \cdot \frac{1}{t_U} \times \frac{2}{(t_1 - t_4)^2}, \; t_L < t_4 < t_3 < t_2 < t_U. \tag{4.8}$$

The marginal prior densities of the calibration node ages (t_1 and t_4) can be obtained by integration.

$$f(t_4) = \int_{t_4}^{t_U} f_C(t_1, t_4)\, \mathrm{d}t_1 = \frac{1}{k_1 t_U}\theta\, \mathrm{e}^{-\theta(t_4 - t_L)} \cdot (t_U - t_4), \quad t_L < t_4 < t_U, \tag{4.9}$$

$$f(t_1) = \int_{t_L}^{t_1} f_C(t_1, t_4)\, \mathrm{d}t_4 = \frac{1}{k_1 t_U}\left[1 - \mathrm{e}^{-\theta(t_1 - t_L)}\right], \quad t_L < t_1 < t_U, \tag{4.10}$$

Note that eq. 4.9 can also be derived by integrating out t_1, t_2, t_3 from $f(t_1, t_2, t_3, t_4)$, and eq. 4.10 can be derived by integrating out t_2, t_3, t_4 from $f(t_1, t_2, t_3, t_4)$. Figure 4.3a (st1) shows the user-specified calibration densities and the effective (marginal) priors after the truncation.

In the *multiplicative* construction used by BEAST and MrBayes, the densities for the calibration nodes of eqs. 4.4 and 4.5 are multiplied with the joint density of the ages of all non-root nodes from the birth-death-sampling process (eq.4.3) to give

$$\begin{aligned} f(t_1,t_2,t_3,t_4) &\neq \tfrac{1}{k_2} \times f_C(t_1) \cdot f_C(t_4) \cdot f_{\mathrm{BDS}}(t_2,t_3,t_4 \mid t_1) \\ &= \tfrac{1}{k_2} \times \tfrac{1}{t_U} \cdot \theta\, \mathrm{e}^{-\theta(t_4 - t_L)} \cdot \tfrac{3!}{t_1^3}, \qquad t_L < t_4 < t_3 < t_2 < t_1, \end{aligned} \tag{4.11}$$

where $k_2 = \int_{t_L}^{t_U}\int_{t_L}^{t_1}\int_{t_L}^{t_2}\int_{t_L}^{t_3} \frac{1}{t_U} \cdot \theta\, \mathrm{e}^{-\theta(t_4 - t_L)} \cdot \frac{3!}{t_1^3}\, \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 \mathrm{d}t_1 = 0.0174371$ is a normalizing constant. Note that eq. 4.11 does not make mathematical sense as two different densities occur for t_4, one in $f_C(t_4)$ and the other in $f_{\mathrm{BDS}}(t_2, t_3, t_4 \mid t_1)$. The marginal (effective) priors for the calibration node ages (t_1 and t_4) can be obtained by integration

$$\begin{aligned} f(t_1) &= \int_{t_L}^{t_1}\int_{t_L}^{t_2}\int_{t_L}^{t_3} f(t_1,t_2,t_3,t_4)\, \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 = \tfrac{3!}{k_2 t_U t_1^3}\left[\tfrac{(t_1 - t_L)^2}{2} + \tfrac{t_L - t_1}{\theta} + \tfrac{1 - e^{-\theta(t_1 - t_L)}}{\theta^2}\right], \\ f(t_4) &= \int_{t_4}^{t_U}\int_{t_3}^{t_U}\int_{t_2}^{t_U} f(t_1,t_2,t_3,t_4)\, \mathrm{d}t_1 \mathrm{d}t_2 \mathrm{d}t_3 = \tfrac{3\theta e^{-\theta(t_4 - t_L)}}{k_2 t_U}\left[\log \tfrac{t_U}{t_4} - 1.5 + \tfrac{2t_4}{t_U} - \tfrac{t_4^2}{2{t_U}^2}\right], \end{aligned} \tag{4.12}$$

with $t_L < t_1 < t_U$ and $t_L < t_4 < t_U$. Figure 4.3b (st1) shows the user-specified calibration densities and the effective (marginal) priors after the truncation.

Calibration strategy 2 (st2). The minimum and maximum bounds are propagated onto all calibration nodes so that the calibration densities are

$$\begin{aligned} f_C(t_1) &= 1/(t_U - t_L), \quad t_L < t_1 < t_U, \\ f_C(t_4) &= 1/(t_U - t_L), \quad t_L < t_4 < t_U. \end{aligned} \tag{4.13}$$

We first consider the *conditional* construction. After truncation, the effective joint prior for t_1 and t_4 becomes, in contrast to eq.4.6,

$$f_c(t_1, t_4) = {}^{2}/_{(t_U - t_U)}, t_L < t_4 < t_1 < t_U. \quad (4.14)$$

This is multiplied with the birth-death-sampling process density for the non-calibration nodes of eq. 4.7 to give the time prior as

$$f(t_1, t_2, t_3, t_4) = f_C(t_1, t_4) \cdot f_{\mathrm{BDS}}(t_2, t_3 \mid t_1, t_4) = \frac{2}{(t_U - t_L)^2} \times \frac{2}{(t_1 - t_4)^2}, \quad t_L < t_4 < t_3 < t_2 < t_1 < t_U.. \quad (4.15)$$

The marginal densities for the calibration node ages are

$$\begin{aligned} f(t_1) &= \int_{t_L}^{t_1} f_C(t_1, t_4)\, \mathrm{d}t_4 = \frac{2}{(t_U - t_L)^2} \times (t_1 - t_L), \quad t_L < t_1 < t_U, \\ f(t_4) &= \int_{t_4}^{t_U} f_C(t_1, t_4)\, \mathrm{d}t_1 = \frac{2}{(t_U - t_L)^2} \times (t_U - t_4), \quad t_L < t_4 < t_U. \end{aligned} \quad (4.16)$$

Figure 4.3a (st2) shows the densities.

With the *multiplicative* construction, the time prior is given by multiplying the calibration densities (eq.4.13) with the birth-death-sampling density for the noncalibration nodes (eq.4.3) and then applying truncation

$$f(t_1, t_2, t_3, t_4) = \frac{1}{k_3} \times \frac{1}{(t_U - t_L)^2} \cdot \frac{3!}{t_1^3}, \; t_L < t_4 < t_3 < t_2 < t_1 < t_U, \quad (4.17)$$

where $k_3 = \int_{t_L}^{t_U} \int_{t_L}^{t_1} \int_{t_L}^{t_2} \int_{t_L}^{t_3} \frac{1}{(t_U - t_L)^2} \cdot \frac{3!}{t_1^3} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 \mathrm{d}t_1$ = 0.00530524 is a normalizing constant, calculated numerically. The marginal (effective) priors for the calibration nodes (t_1 and t_4) are then

$$\begin{aligned} f(t_1) &= \int_{t_L}^{t_1} \int_{t_L}^{t_2} \int_{t_L}^{t_3} f(t_1, t_2, t_3, t_4)\, \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 = \frac{3}{k_3 (t_U - t_L)^2} \left[\log \frac{t_U}{t_4} - 1.5 + \frac{2t_4}{t_U} - \frac{t_4^2}{2t_U^2} \right], \\ f(t_4) &= \int_{t_4}^{t_U} \int_{t_3}^{t_U} \int_{t_2}^{t_U} f(t_1, t_2, t_3, t_4)\, \mathrm{d}t_1 \mathrm{d}t_2 \mathrm{d}t_3 = \frac{(t_1 - t_L)^3}{k_3 t_1^3 (t_U - t_L)^2}, \end{aligned} \quad (4.18)$$

with $t_L < t_1 < t_U$ and $t_L < t_4 < t_U$. Figure 4.3b (st2) shows the user-specified calibration densities and the effective (marginal) priors after the truncation.

Calibration strategy 3 (st3). The minimum and maximum bounds are propagated onto all nodes on the phylogeny, so that every node has joint bounds: $f_C(t_i) = 1/(t_U - t_L)$, $t_L < t_i < t_U$, for $i = 1, 2, 3, 4$. With the *conditional* construction, the birth-death-sampling model plays no role in the construction of the time prior since all nodes have calibration information. After truncation, the effective time prior is

$$f(t_1, t_2, t_3, t_4) = \frac{4!}{(t_U - t_L)^4}, \; t_L < t_4 < t_3 < t_2 < t_1 < t_U, \tag{4.19}$$

Since t_4 is the smallest of four independent and identically distributed (i.i.d.) random variables and t_1 is the largest, their marginal densities are given by the distribution of order statistics

$$\begin{aligned} f(t_1) &= 4 \cdot \left(\frac{t_1 - t_L}{t_U - t_L}\right)^3 \cdot \frac{1}{t_U - t_L}, \quad t_L < t_1 < t_U, \\ f(t_4) &= 4 \cdot \left(\frac{t_U - t_4}{t_U - t_L}\right)^3 \cdot \frac{1}{t_U - t_L}, \quad t_L < t_4 < t_U. \end{aligned} \tag{4.20}$$

Figure 4.3a (st3) shows the densities. Truncation now has a strong effect.

With the *multiplicative* construction, two options seem possible. The first is to ignore the birth-death process density since all the node ages have calibration with this strategy. This is then equivalent to the conditional construction of MCMCTree. The second is to multiply the calibration densities (eq. 4.19) with the birth-death-sampling density of eq. 4.3, followed by a truncation to give

$$f(t_1, t_2, t_3, t_4) = \frac{1}{k_4} \frac{4!}{(t_U - t_L)^4} \times \frac{3!}{t_1^3}, \; t_L < t_4 < t_3 < t_2 < t_1 < t_U, \tag{4.21}$$

where the normalizing constant $k_4 = \int_{t_L}^{t_U} \int_{t_L}^{t_1} \int_{t_L}^{t_2} \int_{t_L}^{t_3} \frac{4!}{(t_U - t_L)^4} \times \frac{3!}{t_1^3} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 \mathrm{d}t_1 =$ 0.00001571922. The marginal priors for t_1 and t_4 are then

$$\begin{aligned} f(t_1) &= \int_{t_L}^{t_1} \int_{t_L}^{t_2} \int_{t_L}^{t_3} f(t_1, t_2, t_3, t_4) \, \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 = \frac{4! \times 3}{k_4 (t_U - t_L)^4} \left[\log \frac{t_U}{t_4} - 1.5 + \frac{2t_4}{t_U} - \frac{2t_4^2}{2t_U^2} \right], \\ f(t_4) &= \int_{t_4}^{t_U} \int_{t_3}^{t_U} \int_{t_2}^{t_U} f(t_1, t_2, t_3, t_4) \, \mathrm{d}t_1 \mathrm{d}t_2 \mathrm{d}t_3 = \frac{4! \times (t_1 - t_L)^3}{k_4 t_1^3 (t_U - t_L)^4}, \end{aligned} \tag{4.22}$$

with $t_L < t_1 < t_U$ and $t_L < t_4 < t_U$. Figure 4.3b (st3) shows the user-specified calibration densities and the effective (marginal) priors after the truncation.

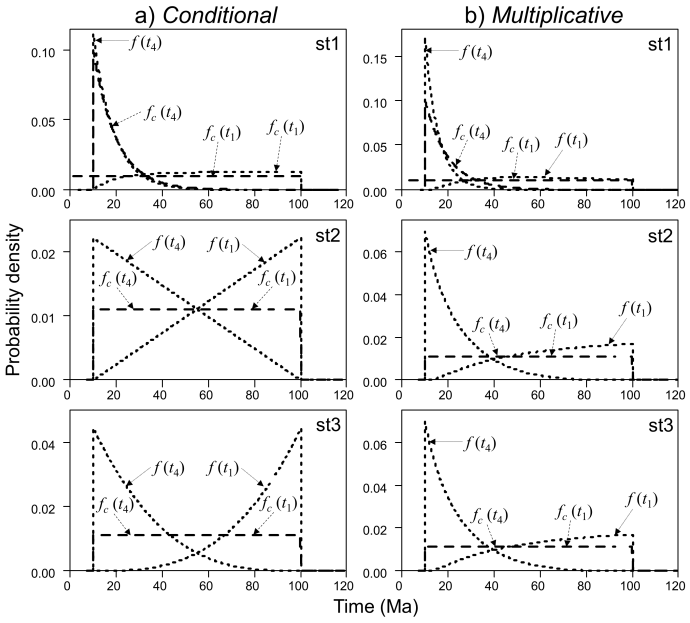


Figure 4.3: User-specified calibration densities and the effective (marginal) priors after the truncation. (a) under the conditional, and **(b)** the multiplicative construction, for st1, st2 and st3.

4.3.4 Analysis of the primate dataset

We used eight mitochondrial coding genes (*Cyt B, CO1, CO2, CO3, ND2, ND3, ND4 and ND4L*) and the mitochondrial 12_S and 16_S ribosomal RNA (rRNA) genes from nine primate species and an outgroup (*Tupaia belangeri*) (Figure 4.4a) (GenBank accession numbers in Table C.3). We partitioned the data into three partitions: (1) 1st and 2nd codon positions; (2) 3rd codon positions and (3) rRNA genes. The final alignment had 9,361 base pairs, with 11.1% of missing data. The data were analyzed using the three dating programs (MCMCTree, BEAST2, and MrBayes), under the independent-rates model to construct the prior of the rates. The time unit is set at 100 Myrs. The sequence likelihood was calculated under the HKY+Γ_5 substitution model (Hasegawa et al., 1985; Yang 1994), with separate rates and exchange-rate parameters for each partition.

There are nine fossil calibrations on the tree (Figure 4.4a) (dos Reis et al., 2012), five of which are joint minimum and maximum bounds, while the other four are minimum

bounds only (Table C.1). We implemented calibration strategies st1 and st2 in the programs MCMCTree, BEAST2, and MrBayes. As all nine interior nodes have calibration information, st3 is equivalent to st2. Bounds are soft in MCMCTree, and hard in BEAST2 and MrBayes. Minimum bounds are implemented using the truncated Cauchy distribution in MCMCTree and the offset-exponential distribution in BEAST2 and MrBayes.

In MCMCTree, the parameters of the birth-death-sampling process are fixed at $\lambda = \mu = 1$, and $\rho = 0$. These specify a uniform kernel. The approximate likelihood method (Thorne et al., 1998; dos Reis & Yang 2011) is used to calculate the sequence likelihood, using the maximum likelihood estimates of branch lengths and the Hessian matrix. The independent-rates model (IR) assumes that the rates for branches are independent variables from the lognormal distribution, specified by the mean of the rate (μ) and the variance of the log rate σ^2 (which determines the extent of rate variation across branches) (Rannala & Yang 2007). The mean rate is assigned a gamma prior $\mu \sim G(2, 2)$ with mean 2/2 = 1.0 substitutions per site per time unit (100MY) or 10^{-8} substitutions per site per year, and the rate drift parameter is assigned another gamma prior, $\sigma^2 \sim G(1,10)$, with mean 0.1.

Both BEAST2 and MrBayes assign hyperpriors to implement the birth-death-sampling model: the net diversification rate $\lambda - \mu \sim U(0, 1)$ and the relative extinction rate $\mu/\lambda \sim U(0, 1)$ (Stadler 2010; Hohna et al., 2011). In MrBayes the sampling probability (ρ) is fixed at 0.02.

In BEAST2 we specified a *Relaxed Clock Log Normal* (ucld) model, which assumes that the substitution rates for branches are independent variables from a lognormal distribution (Drummond et al., 2006). The lognormal distribution is parametrized using the mean and the standard deviation. The mean (*ucldMean.c*) was assigned a gamma hyperprior G(2, 0.5) with mean 1.0, and the standard deviation (*ucldStdev.c*) was assigned a gamma hyperprior G(2, 0.05) with mean 0.1.

In MrBayes we used the *Independent Gamma Rate* (IGR) model in where the rates for branches are independent variables from a gamma distribution (Lepage et al., 2007). The gamma model is parametrized using two parameters: the mean and

variance. The mean is assigned a lognormal hyperprior LN(−0.125, 0.5), with the mean $\exp\{-0.125 + 0.5^2/2\} = 1.0$. The variance (*lgrvarpr*) is assigned an exponential prior with mean 0.1

The MCMC sampling settings were determined through pilot runs and differed among the programs. We ran each program at least twice, and checked for convergence by comparing the posterior mean estimates between runs and by plotting the time series traces of the samples. We then merged the samples from the runs before summarizing the posterior. For MCMCTree, two runs were performed, each consisting of 2×10^6 iterations after a burn-in of 4×10^4 iterations and sampling every 200, resulting in a total of 2×10^4 samples from the two runs. For MrBayes, two runs were performed, each consisting of 2×10^6 iterations, sampling every 100, with the burn-in set to 25% of samples, resulting in a total of 3×10^4 samples from the two runs. For BEAST2 we performed three runs, each consisting of 10^7 iterations, sampling every 1000. The burn-in was set to 30% of samples, resulting in a total of 21,000 samples from all three runs.

4.3.5 Analysis of seed plant dataset

We used five plastid genes (*atpB, matK, NdhF, rbcL*, and *rps4*) and two nuclear RNA genes (*18s* and *26s*) for 48 seed plant species (GenBank accession numbers in Table C.4) from (Barba-Montoya et al., Submitted 2017). The tree topology of Figure 4.4b is fixed. The sequence alignment had 13,211 base pairs, with 26% missing data. We treated the data as three partitions: (1) 1^{st} and 2^{nd} codon positions for plastid genes; (2) 3^{rd} positions for plastid genes and (3) nuclear RNA genes. The data were analyzed using the three programs (MCMCTree, BEAST2, and MrBayes), with similar settings as in the analysis of the primate dataset, but some modifications were necessary to accommodate the differences in the time scale and in the rate. The sequence likelihood was calculated under the HKY+Γ_5 substitution model (Hasegawa et al., 1985; Yang 1994), with separate rates and exchange-rate parameters for each partition. The phylogenetic likelihood is calculated approximately in MCMCTree (Thorne et al., 1998; dos Reis & Yang 2011) and exactly in BEAST and MrBayes.

There are 15 fossil calibrations on the tree (Figure 4.4b) (Barba-Montoya et al., Submitted 2017). Among them seven are joint minimum and maximum bounds and

eight are minimum bounds (Table C.2). The time unit is set to 100 Myrs. The calibration information is implemented in the three programs using the three strategies as described earlier.

In MCMCTree, the parameters of the birth-death-sampling process are fixed at $\lambda = \mu = 1$, and $\rho = 0$. These specify a uniform kernel. The independent-rates model (IR) assumes that the rates for branches are independent variables from the lognormal distribution, with the mean of the distribution (μ) assigned a gamma hyperprior $\mu \sim$ G(2, 30) with mean 2/30 = 0.067 substitutions per site per 100MY or 6.7×10^{-10} substitutions per site per year, and with the variance of the log-rate assigned a gamma hyperprior $\sigma^2 \sim$ G(2, 20) with mean 0.1. Two runs were performed, each consisting of 10^6 iterations after a burn-in of 40,000 iterations and sampling every 200. The combined sample of 10,000 samples was used to summarize.

In the BEAST2 and MrBayes analyses, hyperpriors are assigned to parameters in the birth-death-sampling model: $\lambda - \mu \sim$ U(0, 1) and $\mu/\lambda \sim$ U(0, 1) (Stadler 2010; Hohna et al., 2011). In MrBayes, the sampling probability (ρ) is fixed at 0.0002.

In BEAST2 we specified the ucld model, which assumes that the substitution rates for branches are independent variables from a lognormal distribution. The mean of the lognormal (*ucldMean.c*) was assigned a gamma hyperprior G(2, 0.0335) with mean 0.067, and the standard deviation of the lognormal (*ucldStdev.c*) was assigned a gamma hyperprior G(2, 0.05) with mean 0.1. Three runs were performed, each consisting of 10^7 iterations, sampling every 1000. The burn-in was set to 30% of samples, resulting in a total of 21,000 samples from the posterior from the three runs. In MrBayes we used the *Independent Gamma Rate* (IGR) model. The mean of the gamma was assigned a lognormal hyperprior LN(–2.79, 0.5^2), with the mean $\exp\{-2.79 + 0.5^2/2\} = 0.07$, and the variance of the gamma is assigned an exponential hyperprior with mean 0.1. Four runs were performed, each consisting of 1.5×10^6 iterations, sampling every 100. The burn-in was set to 33.3% of samples, resulting in a total of 4×10^4 samples from all four runs.

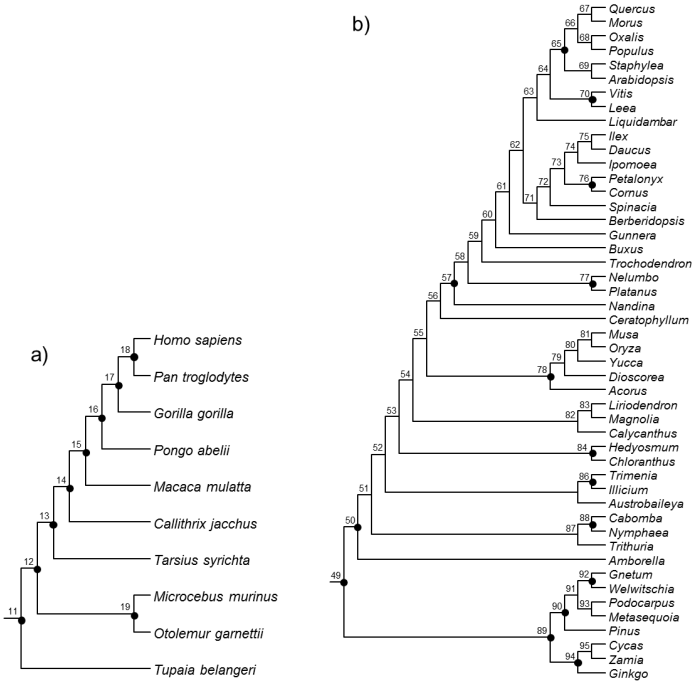


Figure 4.4: Phylogenies for a) 10 primate species, and b) 48 seed plant species. Calibration nodes are indicated by black solid circles.

4.4 Results

4.4.1 Analysis of a simple example with five species

The calibration densities and the effective time priors generated by the conditional and the multiplicative constructions using the three calibration strategies are plotted in Figure 4.3. From Figure 4.3a it is apparent that with the conditional construction strategy st1 generates marginal priors that are closest to the original calibration densities. This is because the youngest node is calibrated with an offset-exponential distribution with a relatively short tail, and so truncation between the two calibration densities is minimal. In Strategy st2 the youngest node inherits the maximum age constraint from the root. This strategy avoids the choice of arbitrary parameters in the Cauchy or shifted-exponential calibrations. In this case truncation is more severe, and the marginal prior densities differ substantially from the calibration densities. In strategy st3, the inclusion of two additional calibration densities for t_3 and t_2 increases the truncation effect, and the result is that the marginal priors on t_4 and t_1 are pushed apart.

The multiplicative construction is shown in Figure 4.3b. Strategy st1 generates marginal priors that are closest to the original calibration densities, while truncation has a major impact in strategies st2 and st3, so that the marginal prior densities differ substantially from the calibration. St2 and st3 generate nearly identical prior densities. Overall Figure 4.3 shows that the conditional and the multiplicative constructions, as well as the different calibration strategies, generate very different effective time priors.

Figure 4.5 shows the results from analysing this example using the three different dating programs. In MCMCTree (Figure 4.5a) the calibration density used for t_4 in strategy st1, is the Cauchy distribution (shifted-exponential) with parameters t_L =10, p = 0.2, c = 0.5 and p_L = 0.0001. We fix the parameters in the birth-death-sampling model at $\lambda = \mu = 1$ and $\rho = 0$ in all strategies. The prior densities generated by the three calibration strategies using MCMCTree (Figure 4.5a, st1, st2, st3) are almost identical to those from the conditional construction (Figure 4.3a, st1, st2, st3).

To examine the implementation in MrBayes and BEAST (Figures 4.5b, c, and d) we fix the parameters in the birth-death-sampling model at $\lambda = \mu = 1$ and $\rho = 0$. To avoid numerical problems, we used λ= 1.001, μ = 0.999 and ρ = 0.0001. In MrBayes the net diversification rate $\lambda - \mu$ is fixed at 0.002, the relative extinction rate μ / λ is fixed at 0.998 and the sampling probability (ρ) is fixed at 0.0001. In BEAST1 and BEAST2 we use for the net diversification rate $\lambda - \mu$ a uniform distribution U(0.00199,0.00201) and for the relative extinction rate μ / λ U(0.99799,0.99801). In BEAST1 we use U(0.000099,0.000101) for the sampling probability (ρ). None of these programs generated identical results to the multiplicative construction. The prior densities generated by MrBayes and BEAST1 were similar but not identical. Precise reasons for the discrepancies between the analytical example, BEAST1 and MrBayes are unknown. One possible reason is that BEAST1 and MrBayes may not be conditioning the birth-death-sampling age density on both root (t_1) or on N only. Nevertheless, based on Figure 4.5, we emphasize the large differences in the prior generated by the conditional and multiplicative constructions and from the three calibration strategies.

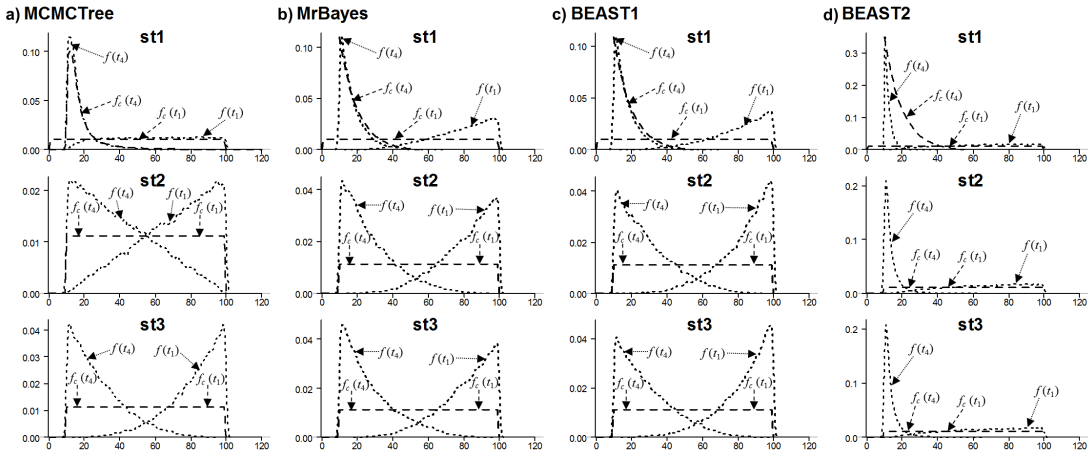


Figure 4.5: User-specified calibrations and effective priors for node ages t_1 and t_4 under three calibration strategies (st1, st2, st3). In a simple example of five species (Figure 4.1), generated using **(a)** MCMCTree; b) MrBayes; **(c)** BEAST1 and **(d)** BEAST2. Dashed lines represent the user-specified calibration densities, while dotted lines represent the effective prior densities.

4.4.2 Analysis of the primate dataset

The calibration densities and the effective time priors generated by the three programs using calibration strategies st1 and st2 are plotted in Figures 4.6 and 4.7. The posterior distributions of divergence times are shown in Figures 4.7 and 4.8.

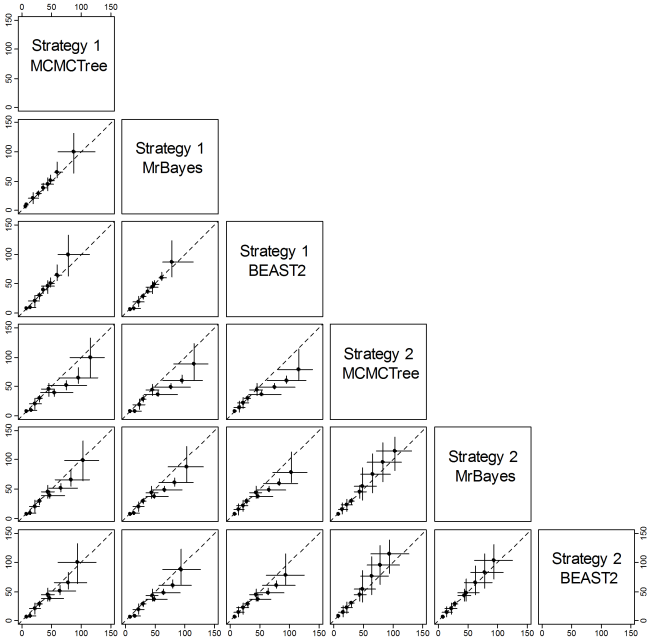


Figure 4.6: Means and 95% CIs in the time prior for node ages on the primate phylogeny (Figure 4.4a). Generated using calibration strategies st1 and st2 and three dating programs: MCMCTree, BEAST2 and MrBayes.

First, we note that with both st1 and st2, the user-specified calibration densities are very different from the marginal densities for the node ages in the effective time prior. This difference is mainly caused by the truncation to enforce the constraint that ancestors are older than descendants. In particular, the root age assigned a pair of bounds represented by the uniform distribution, and in the time prior, the density is pushed towards the maximum. Node 18 is a descendent to many other interior nodes but is ancestral to none, so that its density is pushed towards the minimum. The patterns for other nodes are more complex. Second, strategy st2, which uses uniform bounds for all interior nodes, show much greater truncation effect so that the user-specified calibration densities and the marginal prior densities are even more different than under strategy st1. Third, the differences in the prior of node ages are transferred to the differences in the posterior. For example, the prior favoured much older age for the root under st2 than under st1 for all three programs (Figure 4.7a, b, c, node 11), and this pattern persisted in the posterior.

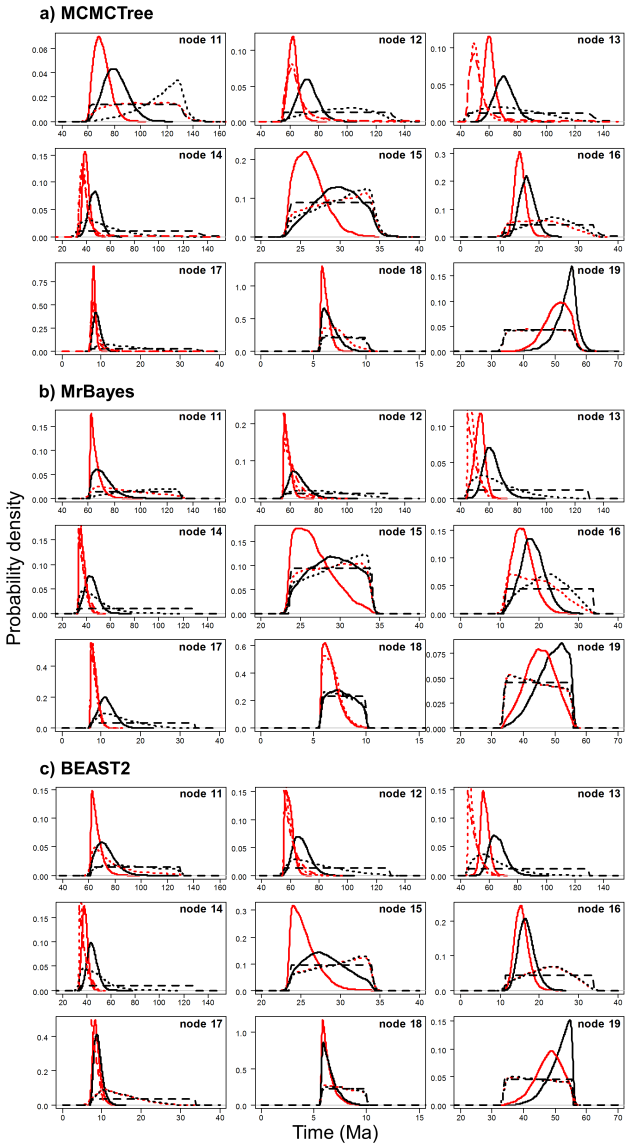


Figure 4.7: User-specified calibration densities (dashed lines), effective time priors (dotted lines), and the posterior (solid lines) for the primate dataset. Under calibration strategies st1 (red) and st2 (black), implemented in **(a)** MCMCTree, **(b)** MrBayes and **(c)** BEAST2.

Lastly, the three dating programs produced similar priors and posteriors (Figures 4.7 and 4.8), although MCMCTree produced slightly older time estimates and wider intervals, especially for old nodes such as the root.

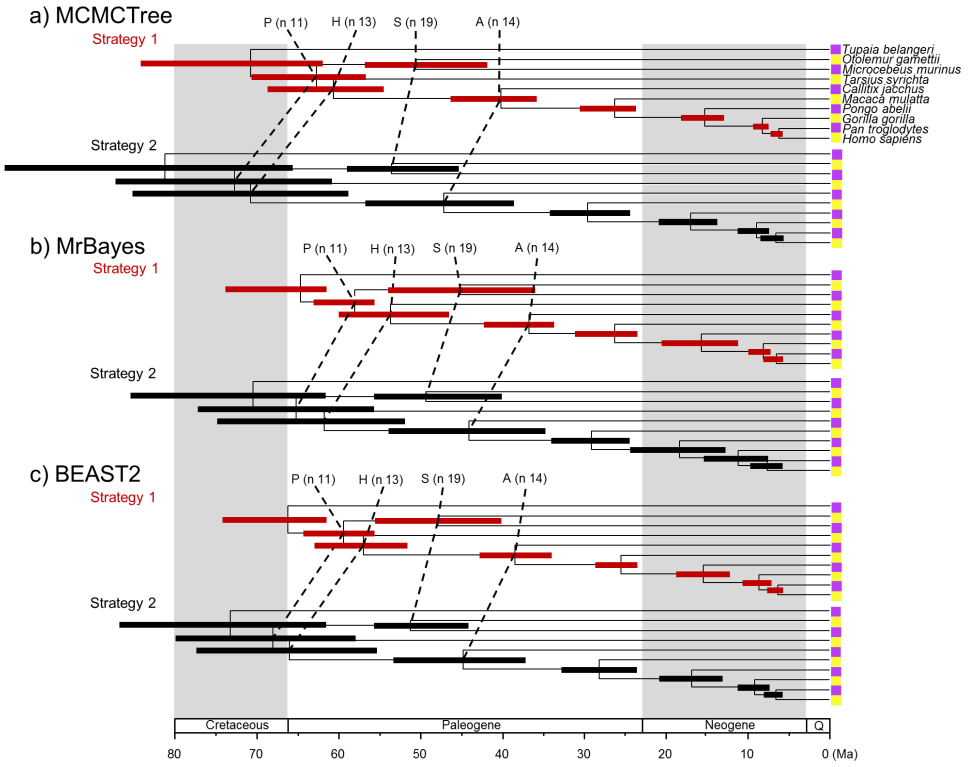


Figure 4.8: Timetrees showing posterior divergence time estimates for the primates.. The branches are drawn to reflect the posterior means of node ages and the bars represent 95% HPD intervals. The dataset was analysed using **(a)** MCMCTree, **(b)** MrBayes and **(c)** BEAST2 under the independent-rates model, using calibration strategies st1 and st2.

4.4.3 Analysis of seed plant dataset

The calibration densities and the effective time priors generated by the three programs using the three calibration strategies are plotted in Figures 4.9 and 4.10. The posterior distributions of node ages are shown in Figures 4.10 and 4.11. We see similar patterns to those in the analysis of the primate dataset. First, there are large differences between calibration densities specified by the user on one hand and the (marginal) effective prior densities used by the dating software on the other. The difference is particularly pronounced for nodes with wide joint bounds as the effective prior used by the dating software is much narrower. Furthermore, truncation pushes the ages of old nodes such as the root towards the user-specified maximum bound, or even outside the maximum bound in the case of MCMCTree which allows bound violation due to its use of soft bounds (e.g., Figure 4.10a, b, and c, node 49). At the

same time, truncation has the effect of pushing the ages of younger nodes towards the minimum bound in the prior (e.g., Figure 4.10a, b, and c, nodes 86, 88, and 89).

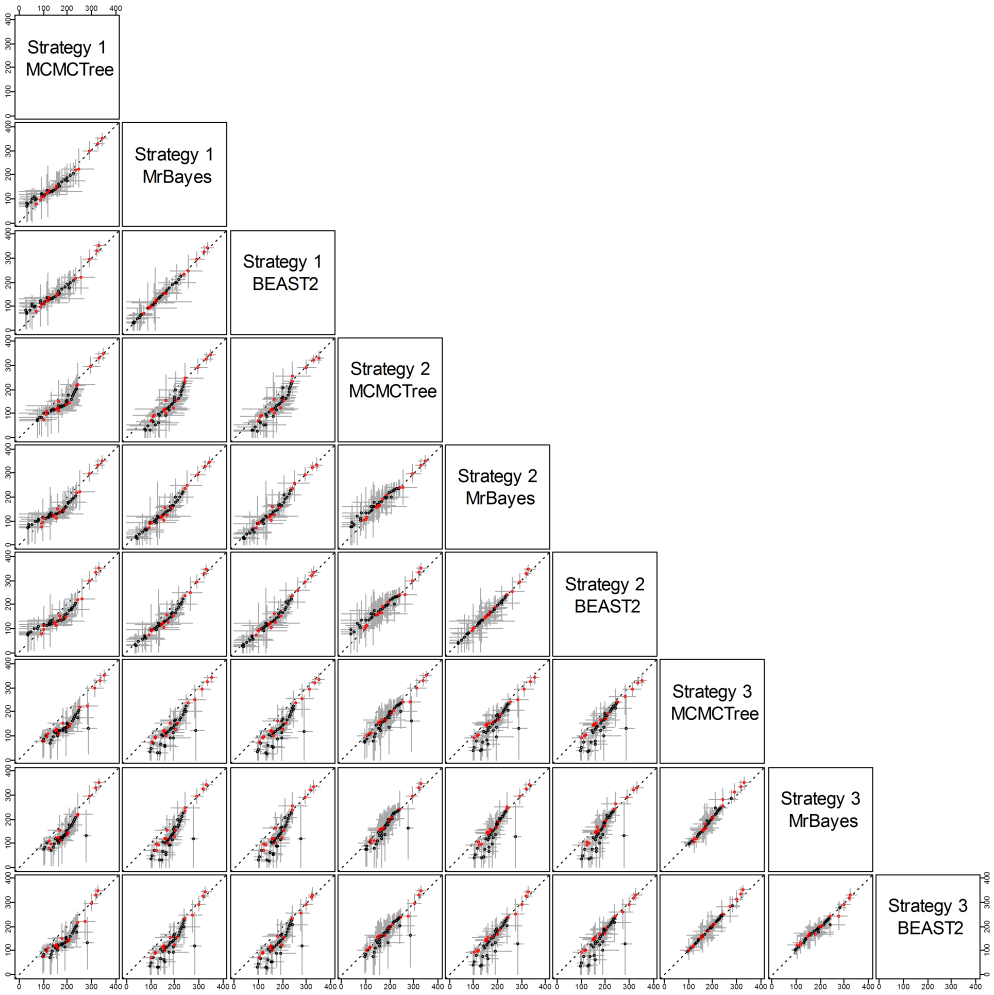


Figure 4.9: Means and 95% CIs in the time prior for node ages on the seed plant phylogeny (Figure 4.4b). Generated using three calibration strategies (st1-3) and three dating programs: MCMCTree, BEAST2 and MrBayes. Calibration nodes are highlighted in red.

Second, as in the case of the primate dataset, the posterior of the node ages is sensitive to the prior, and differences in the time prior are directly transferred to differences in the posterior. For example, nodes 77 and 78 are older under st2 than under st1 and even older under st3, and exactly the same trend persists in the posterior (Figure 4.10a, b, c). This pattern holds for all three dating programs.

Third, strategies st2 and st3 showed greater truncation effects so that the user-specified calibration densities and the marginal prior densities are even more different

than under st1. The large differences in the priors of the three strategies persisted in the posterior. The time estimates tended to be older under st2 than under st1, while st3 produced the oldest time estimates (Figures 4.10 and 4.11). For example, the posterior mean estimated using st1 suggests that the eudicots (node 57) originated around 155 Ma, but using st3 the posterior mean was around 195 Ma, with a difference of 40 Myrs. The origin of monocots (node 78) was dated to ~136 Ma under st1 in BEAST2 and MrBayes and 150 Ma in MCMCTree, but using st3 the posterior mean for this node was around 190 Ma, with again a difference of ~40 Myrs. These differences in the posterior reflect the differences in the time prior generated under the three strategies (Figures 4.10 and 4.11).

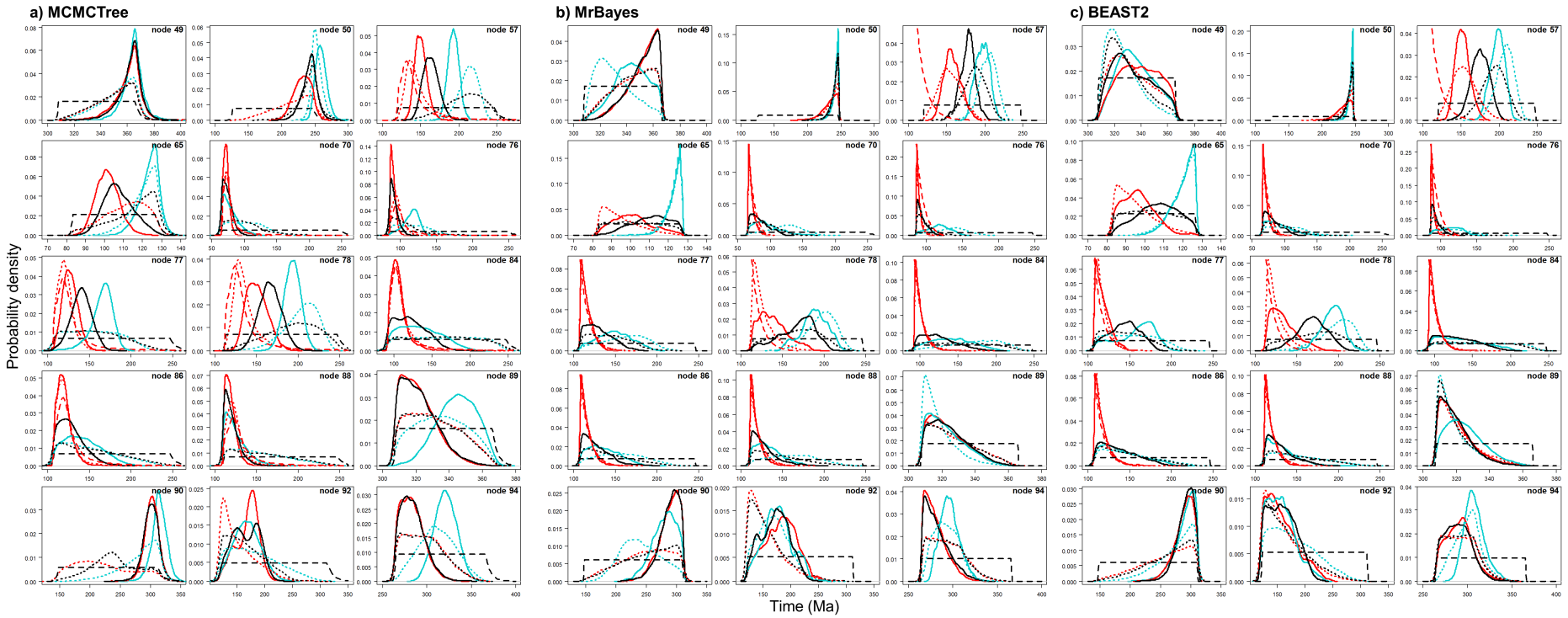


Figure 4.10: User-specified calibration densities (dashed lines), effective time priors (dotted lines), and the posterior (solid lines) for the seed plant dataset. Under calibration strategies st1 (red), st2 (black), and st3 (blue), implemented in **(a)** MCMCTree, **(b)** MrBayes and **(c)** BEAST2. Only the 15 calibration nodes are used in the plots.

Differences in posterior time estimates exist among the three dating programs, reflecting their different procedures to construct the time prior using the same fossil-calibration information (Figures 4.10 and 4.11). BEAST2 produced slightly younger estimates of root age (node 49) and MCMCTree produced narrower intervals than BEAST2 and MrBayes. The differences among the dating programs in both the prior and the posterior are the smallest for calibration strategy st3. This is because with st3 all nodes on the phylogeny were calibrated, so that the birth-death-sampling process plays no or little role in specifying the time prior.

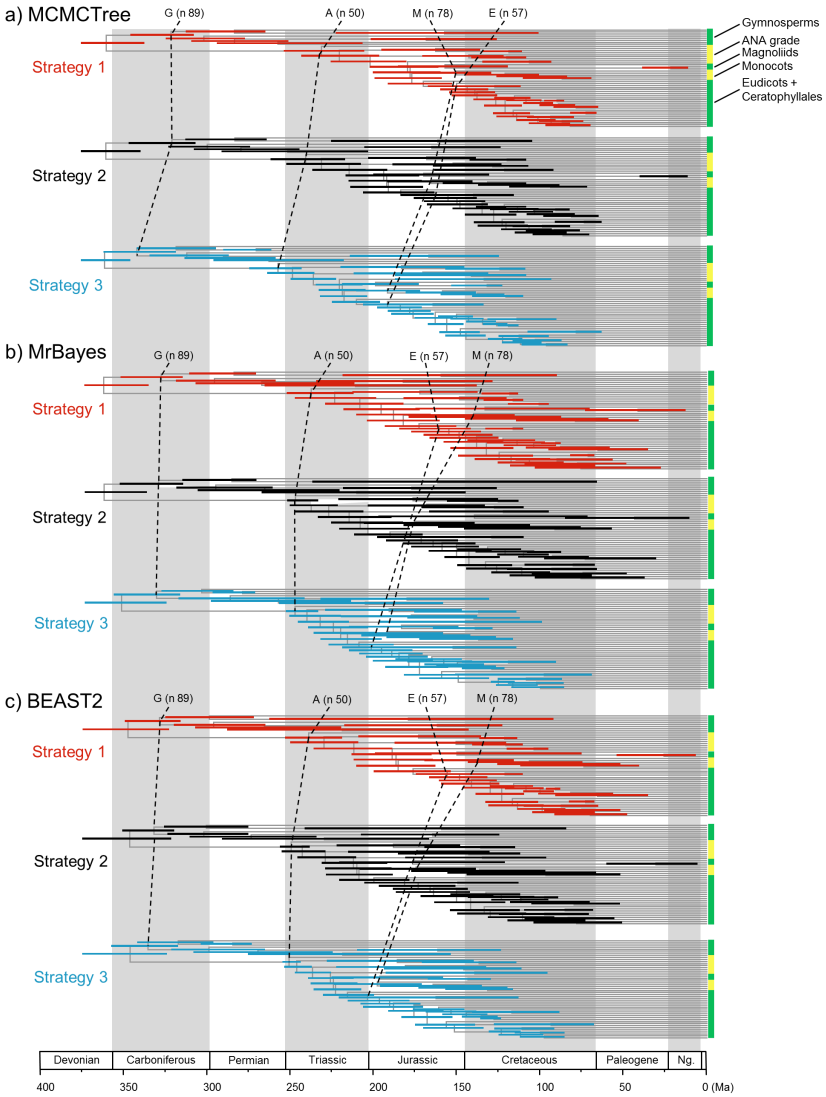


Figure 4.11: Timetrees showing posterior divergence time estimates for major seed plant groups. The branches are drawn to reflect the posterior means of node ages and the bars represent 95% HPD intervals. The dataset was analysed using **(a)** MCMCTree, **(b)** MrBayes and **(c)** BEAST2 under the independent-rates model, using three calibration strategies: st1, st2, and st3.

4.5 Discussion

In a conventional Bayesian analysis, the posterior distribution of the parameters converge to a point mass (the true value of the parameter) and the prior becomes less and less important when the amount of data approaches infinity. Bayesian molecular clock dating is an unconventional estimation problem in the sense that such convergence to truth does not occur (Yang & Rannala 2006). If the amount of molecular data increases and the fossil calibration information is fixed, the posterior will not converge to a point or to the true node ages, and furthermore the prior will continue to exert a large impact on the posterior. Even if we use whole genomes in the dating analysis so that sequence distances and branch lengths are estimated with virtually no random sampling errors, fossil calibrations and the time prior constructed using the fossil calibrations will remain important to the posterior time estimates. The fundamental difficulty faced by the dating analysis is the confounding effect of time and rate in sequence comparisons: molecular data provide information about the genetic distances, and only fossil calibrations (or dated geological events) can resolve the distances into absolute times and absolute rates. The asymptotic dynamics of the dating problem has been characterized in the infinite-sites theory (Yang & Rannala 2006; Rannala & Yang 2007; dos Reis & Yang 2013; Zhu et al., 2015).

Our analyses highlight the arbitrary nature of the procedure used by dating software to generate the time prior from the same fossil calibration information. In the future, we see probabilistic modeling and statistical analysis of fossil data (including both fossil presence/absence data and morphological measurements) as an important approach to summarizing the fossil evidence to generate distributions of divergence times for use as molecular clock calibrations (Tavaré et al., 2002; Wilkinson et al., 2011; Ronquist et al., 2012a; Bracken-Grissom et al., 2014; Heath et al., 2014) For the present, we suggest that the palaeontologist should take a proactive role in constructing calibration densities, by making subjective judgments regarding the quality of the fossil and its placement on the phylogeny. We also encourage the use of the error probabilities in soft-bound calibrations as an approach to represent the uncertainties in the soft maximum bounds. It should be stressed that decisions will be made arbitrarily by the computer program if not subjectively by the palaeontologist. Given that in many cases the resulting time prior can be quite counterintuitively different to the calibration densities, we cannot emphasize enough how important it is

for the user to explicitly calculate the time prior by running the MCMC analysis without data.

In this paper, we have focused on divergence time estimation when fossil calibration information is available on certain nodes on the tree, a procedure called *node calibration*. Recently *tip-calibration* methods have been developed, which analyze fossil data jointly with molecular data, in the so-called fossilized birth-death process model (Heath et al., 2014; Zhang et al., 2016). Morphological characters for both extant and extinct (fossil) species can be incorporated into a joint analysis with the molecular data for extant species (Ronquist et al., 2012a; O'Reilly et al., 2016). The dates for the fossil species provide the calibration information that resolves the morphological distances into absolute times and rates, which are propagated to the other nodes on the phylogeny represented by the molecular data. While the approach shows great promise, it has its own set of challenges (dos Reis et al., 2016; Ronquist et al., 2016). First, morphological characters, driven by natural selection and adaptation to environment and occasionally undergoing convergent evolution, rarely evolving in a clock-like fashion (Kimura 1983). Second, morphological characters may be strongly correlated. Thus, current models (Lewis 2011), which ignore the correlation, are overstating the information content in the data. Third, without constraints on the interior nodes, the Bayesian dating analysis tends to be very sensitive to the birth-death-sampling process used to specify the time prior. Changing the parameters in the branching process may change the shape of the tree (reflected in the relative of internal versus external branch lengths), leading to drastically different posterior time estimates (Drummond & Stadler 2016; Ronquist et al., 2016; Zhang et al., 2016). We believe that both node calibrations and tip calibrations will be used in the foreseeable future (O'Reilly et al., 2015).

5 General conclusions

5.1 The timeline of angiosperm evolution

The timing of origination of Angiosperms is one of the emblematic clock-rock controversies. Previous molecular clock studies involve a number of limitations such as insufficient outgroup lineages, limited sequence data, failure to control for phylogenetic uncertainty and low taxon sampling. However, our attempts to control for them have resulted in greater uncertainty in the molecular timescale, itself allowing for closer agreement, especially for the diversification of clades within crown angiosperms. The time of origin of the crown angiosperm diversification remains in disagreement with traditional interpretations of the fossil evidence, by at least 23 Myr, and as much as 121 Myr. The discovery of equivocal microfossil and macrofossil angiosperm remains in the Jurassic and even Triassic, which together with evidence of biases in the rock record, suggest that this mismatch may be real. Nevertheless, it suggests as well that agreement between clock and rock estimates may be found in quantification of the artifact imposed by the rock record on the fossil record, and the further analysis of equivocal pre-Cretaceous remains. Although the extreme interpretation of a KTR seems incorrect, the diversification of species-rich clades among crown angiosperms does appear to coincide with that of herbivores and pollinators, corroborating a more general hypothesis of a KTR. This highlights the power of the complementary nature of molecular and palaeontological data and approaches to inferring evolutionary timescales and establishing a deeper understanding of clade dynamics in deep time.

This study, jointly with the equivocal pre-Cretaceous crown angiosperm fossils, raises the possibility that the oldest crown angiosperm fossils are still undiscovered or unidentified. However, it is required to reconsider the phylogenetic position of critical fossil taxa using likelihood models of character evolution to accommodate phylogenetic affinity, since discriminating between a stem and crown angiosperm affinity of all pre-Cretaceous claims may be the only way in which molecular estimates for the origin of angiosperms will be more precise.

The angiosperm date estimated in this study cannot be easily rejected as an artifact, as the date is compatible with the current state of knowledge of fossils and phylogeny, and with the current state of development of relevant analytical tools. We believe that these dates should not be lightly discarded. Despite this, it is important to recognize that this study is unlikely to be the final word on the subject.

As taxon sampling increases, and as knowledge of fossils grows, there may be a general trend towards placing fossils further up within their associated lineages. This process will tend to push back in time the age of the angiosperms. It is likely that a closer match between molecular estimates and the fossil record will be eventually achieved, as older fossils are discovered and as dating methods are improved to deal with high rate heterogeneity.

This study will serve as a community resource to further our understanding of angiosperm evolution and diversification. This timescale provides 52 fossil calibrations — including 49 newly formulated calibrations, all of which achieve the expectations of best practice — for future studies of land plant divergence. Furthermore, this timeline represents an ideal framework to investigate, for example, potential hypothesis concerning codiversification with other organisms, rates of molecular evolution, biogeographical history, diversification dynamics, ancestral character reconstruction and state-dependent diversification, correlations between diversification and the physical environment, and the evolution of modern terrestrial ecosystems.

5.2 Using fossil calibrations to generate the time prior in Bayesian molecular clock dating

Fossil calibrations are the major source of information for resolving the distances between molecular sequences into estimates of absolute times and absolute rates in molecular clock dating analysis. The quality of the calibrations is thus expected to have a major impact on divergence time estimates even if a huge amount of molecular data is available. Calibrations can be strongly affected by truncation, so that the effective priors on the calibration node ages after the truncation can be very different from the user-specified calibration densities. The different strategies for generating the effective prior also had considerable impact, leading to very different marginal

effective priors. This study highlights the arbitrary nature of the procedure used by dating software to generate the time prior from the same fossil calibration information, and emphasizes the importance of reporting the time prior in divergence time studies. Furthermore, not all Bayesian divergence time programs construct the time prior in the same way and reporting the time prior is necessary for valid comparisons among different studies.

The development of probabilistic modeling and statistical data analysis of fossil data — including measurements of both fossil presence/absence and morphological data — will be an important approach to summarizing the fossil evidence to generate distributions of divergence times for use as molecular clock calibrations. For the time being, the palaeontologist should concentrate their efforts to construct calibration densities, by making subjective judgments regarding the quality of the fossil and its placement on the phylogeny. We also encourage the use of the error probabilities in soft-bound calibrations as an approach to represent the uncertainties in the soft maximum bounds.

This study is focussed on node calibration. Recently tip-calibration methods have been developed, which analyze fossil data jointly with molecular data, in the so-called fossilized birth-death process model. Moreover, morphological characters for both extant and extinct (fossil) species are analysed under a morphological clock, jointly with the molecular data for extant species. The dates for the fossil species provide the calibration information that resolves the morphological distances into absolute times and rates, which are propagated to the other nodes on the phylogeny represented by the molecular data. Although the latest has its own set of complications it is very likely that in the near future node calibrations and tip calibrations will be used.

5.3 Perspectives and future directions of molecular clock dating of angiosperms divergence times

The results presented in this thesis highlight two issues that should be considered as the field moves forward: the random, non-uniform nature of the rock and fossil records and the adequacy of current relaxed clock models. In this case, the only way forward, given the methods of inference may be to further develop four lines of research. The

first is the development of palaeobotanical research; it is required to reconsider the phylogenetic position of critical fossil taxa using likelihood models of character evolution to accommodate phylogenetic affinity (O'Reilly et al., 2016). Moreover, it is likely that in the future, plants on the angiosperm stem-lineage, with some, but not all of the critical features of extant angiosperms, will be recognized in Jurassic or earlier rocks. It is also possible a significant discovery of a pre-Cretaceous fossil with features that allow it to be assigned with high confidence to the angiosperm stem group or crown group. A discovery like this would probably be enlightening about many aspects of early angiosperms that remain poorly understood. The second is the development of new methods of divergence time inference that largely avoid the concerns related to node-dating. For example, the fossilized birth-death model (Heath et al., 2014; Zhang et al., 2016) that incorporates extant and extinct lineages as evolving according to the same underlying diversification model. The third is the development of methods that can analyse molecular data together with morphological data. When morphological data are available for both extinct and extant taxa, divergence times can be estimated using the total evidence approach (Ronquist et al., 2012a). Finally, the fourth is the improvement of relaxed clock methods that deal with high rate heterogeneity. These two issues should be addressed simultaneously if we want to move forward confidently. I Believe that it is under such as unified framework that the final answers to the origin and timescale of angiosperm evolution will come to light.

Besides the further development of these four lines of research, the analysis of genomic and transcriptomic scale data (e.g. Wickett et al., 2014; Murat et al., 2017) should be considered. Foster et al. (2017) used full chloroplast genomes and demonstrated that the use of many genes had very limited impact on divergence time estimates in angiosperms. However, their study was restricted to 83 plastid genes. Murat et al. (2017) used 1175 nuclear genes and although their dating approach is strange (due to ultralow taxon sampling and the use of only two calibrations) they found essentially the same results as all other recent studies that did not constrain the age of the angiosperms. Wickett et al. (2014) analyzed 852 nuclear genes from 92 taxa of green plants and found consistent results with recent phylogenomic analyses of angiosperm relationships, but they did not perform molecular clock dating. Therefore, it would be interesting to combine (and curate) the nuclear genes from Wickett et al. (2014) and Murat et al. (2017) with the plastid genomes from this study

or others (Ruhfel et al., 2014; Foster et al., 2016) to date the age of the angiosperm divergences. Analyzing genome scale datasets is challenging, but certainly will improve our understanding of angiosperm evolutionary rates and divergence time estimation.

Appendices

A. Justification of fossil calibrations

Node 645 | CG Tracheophyta | MRCA: Lycophyta-Euphyllophyta | 422 Ma – 451 Ma.

Fossil taxon and specimen. Clarke at al. [1] based their calibration of this node on *Zosterophyllum* sp. [US384-8137; University of Saskatchewan Collec but more accurate timescale but more accurate timescale tions, Canada] from Bathurst Island [2].

Phylogenetic justification. Following Clarke et al. (2011), the *Zosterophyllum* sp. from Bathurst Island (Kotyk et al. 2002) is unequivocally zostrophyll given its possession of reniform sporangia, sporangia that dehisce along their distal margins, and laterally inserted sporangia. All *Zosterophyllum* species are total group Lycopsida [3].

Minimum age. 422 Ma.

Maximum age. 451 Ma.

Age justification. *Zosterophyllum* sp. on Bathurst Island [2] co-occurs with conodont *Ozarkodina douroensis*, which is restricted to the Ludlow (as O. n. sp. B in [4-7]. Thus, a minimum age interpretation can be derived from the top of the Ludlow, dated to 423.0 Ma ± 1.0 Myr, thus 422.0 Ma. The maximum constraint, following Clarke et al. [1], is based on the oldest occurrences of trilete spores, known from the Qusaiba-1 core from the Quasim Formation of northern Saudi Arabia [8]. Thus, we establish our maximum for Tracheophyta at 451 Ma. The very oldest records precede the earliest occurrences of the *Acanthochitina barbata*, opening the possibility that they occur in the preceding *Tanuchitina fistulosa* biozone, though *T. fistulosa* does not occur. The oldest stratigraphic records within the core co-occur with the chitinozoan *Armoricochitina nigerica*, known to extend into the Caradoc, to within the biozone characterized by *Fungochitina spinifera* (= *Fungochitina fungiformis*) [9]. The base of the *F. spinifera* zone falls within the *Dicranograptus clingani* Biozone (*Dicellograptus morrisi* subzone) [10], the base of which is estimated at 451 Ma [11].

Node 646 | CG Euphyllophytes | MRCA: Monilophytes-Spermatophyta | 385.571 Ma – 451 Ma.

Fossil taxon and specimen. *Rellimia thomsonii* from the Panther Mountain Formation of New York [12] [335.34; Paleobotanical Collection of the State University of New York at Bingham].

Phylogenetic justification. Magallón et al. [13] identified *Ibyka amphikoma* [14] as the oldest record of the pteridophyte lineage based on phylogenetic analyses undertaken by Kenrick and Crane [3].

Minimum age. 384.71 Ma.

Maximum age. 451 Ma.

Age justification. Clarke et al. [1] proposed *Rellimia thomsonii*, an aneurophytalean progymnosperm from the Panther Mountain Formation of New York [12], as the oldest record of crown Euphyllophyta. The Panther Mountain Formation is equivalent to the Ludlowville and Skaneateles formations [1], which occur below the Moscow Formation of New York [15], making *Rellimia thomsonii* older than *Ibyka amphikoma* [1]. The Ludlowville-Moscow formation boundary falls deep within the Lower *varcus*

zone [16] and, therefore, below the *rhenanus-ansatus* biozonal boundary [17], at the very least, which has been dated to 386.25 Ma ± 0.679 Myr, yielding a minimum constraint of 385.571 Ma. The maximum constraint, following Clarke et al. [1], is based on the oldest occurrences of trilete spores, known from the Qusaiba-1 core from the Quasim Formation of northern Saudi Arabia [8]. The very oldest records precede the earliest occurrences of the *Acanthochitina barbata*, opening the possibility that they occur in the preceding *Tanuchitina fistulosa* biozone, though *T. fistulosa* does not occur. The oldest stratigraphic records within the core co-occur with the chitinozoan *Armoricochitina nigerica*, known to extend into the Caradoc, to within the biozone characterized by *Fungochitina spinifera* (= *Fungochitina fungiformis*) [9]. The base of the *F. spinifera* zone falls within the *Dicranograptus clingani* Biozone (*Dicellograptus morrisi* Subzone) [10] et al. ,2008) [11], the base of which is estimated at 451 Ma (Cooper and Saddler 2012). Thus, we establish our maximum for Tracheophyta at 451 Ma.

Discussion. Magallón et al. [13] established a minimum age constraint using *Ibyka amphikoma,* based on the Givetian-Frasnian boundary, for which they provided a date of 385 Ma, though this has since been revised to 382.7 Ma ± 1 Myr [17]. *Ibyka amphikoma* was recovered from the Manorkill Shale Member, which is a lateral equivalent of the Windom Member, within the Moscow Formation of New York [18, 19], which falls fully within the *ansatus* conodont Biozone [20, 21] the top of which is dated to 385.41 Ma ± 0.7 Myr [17], thus, yielding a minimum age constraint of 384.71 Ma, younger than the minimum age of *Rellimia thomsonii*.

Node 647 | CG Spermatophytes | MRCA: *Ginkgo-Austrobuxus* | 308.14 Ma – 365.629 Ma.

Fossil taxon and specimen. *Cordaites iowensis* [UIC 12,233: University of Illinois at Chicago; OUPH 9616- 9742: Ohio University Paleobotanical Herbarium, Department of Botany, Ohio University, Athens, Ohio] from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa [22].

Phylogenetic justification. Clarke *et al*. [1] identify cordaitean coniferophytes as the oldest records of the crown group of the spermatophyte clade. The oldest whole plant reconstruction is *Cordaites iowensis* from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa [22].

Minimum age. 308.14 Ma.

Maximum age. 365.629 Ma.

Age justification. Janousek and Pope [23] argue that the Laddsdale Coal is equivalent to the Bluejacket Coal of Oklahoma, which occurs as part of the Bluejacket Sandstone Member, underlying the Inola Limestone, part of the Inola Cyclothem of the Krebs subgroup of the Cherokee Group, characterized by the occurrence of the conodonts *Idiognathodus amplificus, Idiognathodus podolskensis and Neognathodus asymmetricus* [24]. The Inola cyclothem falls fully within the *Idiognathodus amplificus/ Idiognathodus obliquus* biozone [25]. This is indicative of the *Neognathodus medexultimus-Streptognathodus concinnus* (Pc10) biozone, certainly older than the *Neognathodus roundyi – Streptognathodus cancellosus* (Pc11) biozone [25, 26]. The base of Pc10 is bracketed by an older age constraint of 312.01 Ma ± 0.37 Myr and the base of Pc11 is bracketed by a younger age constraint of 308.5 Ma ± 0.36 Myr in the Composite Standard of Davydov et al. [26], yielding a minimum constraint of 308.14 Ma.

The maximum constraint follows Clarke et al. [1] who based theirs on the first records of seeds in the form of preovules that satisfy the criteria of the seed habit, which occur in the Upper Fammenian (Late Devonian) VCo Spore Biozone [27], a

well documented example of which being *Elkinsia polymorpha* [28]; *E. polymorpha* has been recovered from the Hampshire Formation, West Virginia, from which the palynomorphs *Grandispora cornuta, Retispora macroreticulata, Retusotriletes phillipsii* and *Rugospora radiata* have been reported [29], which substantiate assignment to the VCo Biozone [30]. The VCo biozone is not directly dated but its base falls within the *Palmatolepis trachytera* conodont biozone [31], the base of which is dated to 364.19 Ma ± 1.439 Myr [17], yielding a maximum constraint on the divergence of crown Spermatophyta at 365.629 Ma.

Node 648 | CG Angiosperms | MRCA: *Amborella-Austrobuxus* | 125.9 Ma – 247.3 Ma.

Fossil taxon and specimen. Tricolpate pollen grain [BRN 126] from the Cowleaze Chine Member of the Vectis Formation of the Isle of Wight [32].

Phylogenetic justification. Following Clarke at al. [1], our minimum age constraint is based on the earliest occurrences Fischer's rule tricolpate pollen, and knowledge of the distribution of tricoplate pollen across the phylogeny of angiosperms [33].

Minimum age. 125.9 Ma.

Maximum age. 247.3 Ma.

Age justification. Following Clarke at al. [1], the Cowleaze Chine Member of the Vectis Formation of the Isle of Wight (Hughes & McDougall, 1990) occurs within the M1n polarity chron at the top of the Barremian, dated as 126.3 Ma ± 0.4 Myr [34]. The maximum age constraint is based on sediments devoid of angiosperm-like pollen below their first report in the Middle Triassic, thus, the base of the Anisian, dated to 247.1 Ma ± 0.2 Myr [35], thus, 247.3 Ma.

Discussion. The recently described *Euanthus panii* [36], *Juraherba bodae* [37] and *Yuhania dahugouensis* [38] from the Jiulongshan Formation were considered but not assigned. At the current stage, the age of the formation appears to be still not fully settled despite most experts agree on a middle Jurassic age (see [37, 38]), whereas the assignment to extant lineages also required further investigation using phylogenetic approaches to confirm the proposed relationships of *Juraherba* to Hydatellaceae - which are the sister to the remaining Nymphaeales lineage (node 1276) and *Yuhania* to monocots.

Node 651 | CG Mesangiosperms | MRCA: *Chloranthus-Austrobuxus* | 125.9 Ma – 247.3 Ma.

Fossil taxon and specimen. Tricolpate pollen grain [BRN 126] from the Cowleaze Chine Member of the Vectis Formation of the Isle of Wight. [32].

Phylogenetic justification. Following Clarke at al. [1], our minimum age constraint is based on the earliest occurrences Fischer's rule tricolpate pollen, and knowledge of the distribution of tricoplate pollen across the phylogeny of angiosperms [33].

Minimum age. 125.9 Ma.

Maximum age. 247.3 Ma.

Age justification. Following Clarke at al. [1], the Cowleaze Chine Member of the Vectis Formation of the Isle of Wight [32] occurs within the M1n polarity chron at the top of the Barremian, dated as 126.3 Ma ± 0.4 Myr [34]. The maximum age constraint is based on sediments devoid of angiosperm-like pollen below their first report in the Middle Triassic, thus, the base of the Anisian, dated to 247.1 Ma ± 0.2 Myr [35], thus, 247.3 Ma.

Node 655 | CG Eudicots | MRCA: *Dicentra-Austrobuxus* | 119.6 Ma.
Fossil taxon and specimen. *Hyrcantha decussata* [NJU-DES02001: Geological Institute, Chinese Academy of Sciences, Beijing], from the lower part of the Yixian Formation, Jehol Group, Liaoning Province, China [39].
Phylogenetic justification. Similar to *Leefrutcus* from the Yixian formation of the Lower Cretaceous of China, *Hycantha* is considered to be a stem group representative of the Ranunculales [40].
Minimum age. 119.6 Ma.
Age justification. The main fossil bearing beds have been dated and may be as old as 129.2 Ma [41], however, in the absence of knowledge of the position of the fossils within the stratigraphy, relative to the sources of the absolute dates, a minimum age constraint can be derived from the Jiufontang Formation which overlies it. $^{40}Ar/^{39}Ar$ dating of a number of samples from the Jiufontang Formation has yielded an age of 120.3 ± 0.7 Ma for the volcanic tuffs [42], establishing a minimum constraint of 119.6 Ma.

Node 701 | SG Malphighia | MRCA: *Dicella-Malphighia* | 44.83 Ma.
Fossil taxon and specimen. *Perisyncolporites pokornyi* [UFP65: Paleobotanical Collection of the Florida Museum of Natural History, University of Florida: Gainesville, Florida, USA] from central Colombia [43].
Phylogenetic justification. Assigned to Malpighiales [43, 44]. Based on morphological similarities, the fossils are considered as reliable assigned to the stem of the stigmaphyllloid clade [45].
Minimum age. 44.83 Ma.
Age justification. The minimum age of *Perisyncolporites pokornyi* is best constrained in sections in central Colombia which Jaramillo and Dilcher [43] integrated into a graphic correlation composite standard and, on this basis, were able to establish its first occurrence as within the Lower Eocene. This composite standard has been refined and calibrated to absolute time by Jaramillo et al. [46] who provide a date of 44.83 Ma for the first appearance of *Perisyncolporites pokornyi.*

Node 753 | SG *Salix* plus *Populus* | MRCA: *Idesia-Salix* | 48.57 Ma.
Fossil taxon and specimen. *Pseudosalix handleyi* [UMNH PB-1: Utah Museum of Natural History,
Salt Lake City, USA] from lacustrine shales of the Parachute Creek Member of the Green River Formation in the vicinity of Bonanza, Utah, USA [47].
Phylogenetic justification. Our node assignment follows the currently accepted interpretation of the fossil record of Salicaceae [48].
Minimum age. 48.57 Ma.
Age justification. The Parachute Creek Member reaches into C22n magnetozone [49], the minimum age of which can be established from the base of the succeeding C21r, dated to 48.57 Ma in the combined age model of Vandenberghe et al. [50].

Node 776 | SG Clusiaceae | MRCA: Clusia-Hypericum | 85.8 Ma.
Fossil taxon and specimen. *Paleoclusia chevalieri* [CUPC 1192: L. H. Bailey Hortorium Paleobotanical Collection, Cornell University, Ithaca, NY, USA] from the Old Crossman locality, New Jersey, USA [51].
Phylogenetic justification. The phylogenetic interpretation follows Ruhfels [52].
Minimum age. 85.8 Ma.
Age justification. Clarke et al. [1] argued that a minimum constraint on the age of this deposit could be established from Santonian-Campanian Boundary, however,

Massoni et al. [53] argue that a tighter correlation can be established with better rocks attributable to the CC13-14 Nannofossil zones in South Carolina, indicating a minimum age of 86.3 Ma ± 0.5 Myr, thus, 85.8 Ma.

Node 825 | CG Fagales | MRCA: *Nothofagus-Fagus* | 85.8 Ma.
Fossil taxon and specimen. *Nothofagidites senectus* [GSV 61898: Mines Department of Victoria, Melbourne, Australia] from the Bass and Gippsland Basins, Australia [54]
Phylogenetic justification. These microfossils are widely considered to be Nothofagaceae which in turn are the sister to the remaining Fagales clade [55].
Minimum age. 85.8 Ma.
Age justification. The first appearance of *Nothofagidites senectus* defines the base of the *Nothofagidites senectus* Palynozone, which coincides with the Coniacian-Santonian Boundary, which is dated to 86.3 Ma ± 0.5 Myr [34], affording a minimum age constraint of 85.8 Ma.

Node 830 | CG Fagaceae | *Fagus-Quercus* | 47.6 Ma.
Fossil taxon and specimen. *Fagus langevinii* [UWBM 97583: Burke Museum, Seattle, Washington, USA] preserved as impressions and carbonaceous films in a siliceous shale in an unnamed formation within the Kamloops Group at the McAbee Locality near the town of Cache Creek, British Columbia [57].
Phylogenetic justification. This fossil is accepted as a stem group representative of the genus *Fagus* [58].
Minimum age. 47.6 Ma.
Age justification. Ashes within the 30 metre sequence of fossiliferous shales have yielded dates based on K-Ar geochronology, ranging between 52 Ma ± 2 Myr and 49 Ma ± 2 Myr [59]. Correlation based on facies and pollen biostratigraphy establish an Ypresian age for the fossil assemblage , compatible with the aged from the intercalated ashes. Therefore, a minimum age constraint can be established on the Ypresian-Lutetian Boundary which has been dated to 47.8 Ma ± 0.2 Myr based on the combined age model of Vandenberghe et al. [50], thus, 47.6 Ma.

Node 848 | SG Polygalaceae | MRCA: *Polygala-Medicago* | 61.6 Ma.
Fossil taxon and specimen. *Paleosecuridaca curisii* [PP34562: Field Museum, Chicago, IL, USA], described by from the Sentinel Butte Formation at the Almont site in Morton County, central North Dakota [62].
Phylogenetic justification. The phylogenetic relationships were clarified by Pigg [61].
Minimum age. 61.6 Ma.
Age justification. The age of the Sentinel Butte Formation has been assigned to the Tiffanian 3 based on mollusk and mammal-based biostratigraphy, but the most concrete age evidence is based on palynostratigraphy, assigned the Almont Site sediments of the Sentinel Butte Formation to Pollen Zone 5 of the late Palaeocene [63]. In the absence of further constraint we establish a minimum age for *Paleosecuridaca curisii* based on the Palaeocene-Eocene boundary, dated to 61.6 Ma in the combined age model of Vandenberghe et al. [50].

Node 875 | CG Sapindales | MRCA: *Citrus-Nitraria* | 59.24 Ma.
Fossil taxon and specimen. *Dipteronia brownii* [UF 15740E-23086: Florida Museum of Natural History, Gainesville FL, USA] from the Paleocene Fort Union Formation at Hell's Half Acre, Wyoming [64].

Phylogenetic justification. This fossil is assigned to the extant genus *Dipteronia* which belongs to the subfamily Hippocantanoides of the family Sapindaceae. The extant genus is considered a Tertiary relict having two extent species endemic to China [65, 66]. Being a possible stem group representative of the extant genus nested in the Sapindales provided the framework for this assignment.
Minimum age. 59.24 Ma.
Age justification. *Dipteronia brownii* occurs within the P4 Pollen Zone in the type section of Nichols and Ott [67], which falls fully within Magnetic Anomaly Zone C26r [68], the end of which is dated to 59.24 Ma in the combined age model of Vandenberghe et al. [50].

Node 878 | SG Ailanthus plus *Citrus* | MRCA: *Ailanthus-Swietenia* | 51.83 Ma.
Fossil taxon and specimen. *Ailanthus confucii* (senior synonym of *Ailanthus lesquereuxi*) [DMNH 7879: Denver Museum of Natural History, Denver CO, USA] from the Fossil Butte fish quarries of the Green River Formation near Kemerrer, Wyoming [69].
Phylogenetic justification. This fossil is accepted to belong to the extant genus *Ailanthus* and can be therefore assigned to this clade [69].
Minimum age. 51.83 Ma.
Age justification. The Fossil Butte Member of the Green River Formation is bounded minimally by the C23r magnetozone [49], the minimum age of which can be established from the base of the succeeding C23n, dated to 51.83 Ma in the combined age model of Vandenberghe et al. [50].

Node 887 | CG Myrtales | MRCA: *Myrtus-Oenothera* | 83.3 Ma.
Fossil taxon and specimen. *Esqueiria futabensis* [PP45419: Field Museum, Chicago IL, USA] from two levels in the Futaba Group exposed in Fukushima Prefecture. northeastern Honshu, Japan [70].
Phylogenetic justification. The phylogenetic relationships have been established by several authors [71].
Minimum age. 83.3 Ma.
Age justification. One locality, considered Coniacian, occurs in the Asamigawa Member of the Ashizawa Formation, on a tributary of the Kitaba River in Kamikitaba, Hirono-machi. Unfortunately, no material evidence has been presented to substantiate this age assignment (Takahashi et al. [70], among others, merely cite the presence of unspecified Coniacian ammonites). The second locality is in the middle part of the Tamayama Formation, on the Kohisa River, Kohisa, Ouhisa machi, northeast of Iwaki City. The Asamigawa Formation is the lowermost formation in the Futaba Group, and is overlain by the Kasamatsu Formation, in turn overlain by the Tamayama Formation. The age of the Tamayama Formation is substantiated on the presence of *Inoceramus amakusensis* [70], which is restricted to the Santonian [72]. Thus, a minimum age constraint may be established on the Santonian-Campanian Boundary, dated as 83.6 Ma ± 0.3 Myr [34], thus, 83.3 Ma.

Node 897 | CG Vitales | MRCA: *Vitis-Leea* | 65.508 Ma.
Fossil taxon and specimen. *Indovitis chitaleyae* [UF19279-56220: Florida Museum of Natural History (UF) Gainesville, Florida, USA] preserved as fruits and seeds in chert from Deccan Intertrappean beds exposed in a quarry near the village of Mahurzari, India about 14 km from Nagpur, along with young fruits with intact seeds and isolated mature seeds from a series of localities including Mohgaonkalan in Chhindwara District and Ambabagholi in Baitul District, both in Madhya Pradesh, and

Shibla in Yeotmal District, Maharashtra [73].
Phylogenetic justification. The phylogenetic relationships of this fossil have been clarified previously [73].
Minimum age. 65.508 Ma
Age justification. A minimum constraint on the age of *Indovitis chitaleyae* can be established on minimum age of Deccan volcanism, which has been constrained to 65.535 Ma ± 0.027 Myr [74], thus, 65.508 Ma.

Node 901 | CG Saxifragales core | MRCA: *Haloragis-Itea* | 85.8 Ma.
Fossil taxon and specimen. *Divisestylus brevistamineus* [CUPC 1340: L. H. Bailey Hortorium Paleobotanical Collection, Cornell University, Ithaca NY, USA], flowers described from the Old Crossman locality, New Jersey, USA [75].
Phylogenetic justification. The phylogenetic relationships of this fossil have been clarified previously [75].
Minimum age. 85.8 Ma
Age justification. Clarke et al. [1] argued that a minimum constraint on the age of this deposit could be established from Santonian-Campanian Boundary, however, Massoni et al. [53] argue that a tighter correlation can be established with better rocks attributable to the CC13-14 Nannofossil zones in South Carolina, indicating a minimum age of 86.3 Ma ± 0.5 Myr, thus, 85.8 Ma.

Node 914 | SG Hamamelidaceae | MRCA: *Daphniphyllum-Rhodoleia* | 82.00 Ma.
Fossil taxon and specimen. *Androdecidua endressii* [PP45947: Field Museum, Chicago IL, USA] from the Buffalo Creek Member of the Gaillard Formation in the south pit (Allon Quarry) of the Atlanta Sand and Supply Company in Gaillard, Georgia, ca. 9.5 km southeast of Roberta in Crawford County [76].
Phylogenetic justification. The phylogenetic relationships were carefully discussed by the authors introducing the taxon [76].
Minimum age. 82.00 Ma
Age justification. Christopher correlated this deposit to nannofossil zone CC17 based on its palynoflora, though the evidence on which this is based was not presented. CC17 ranges in age from Late Santonian to Early Campanian and the CC17-18 boundary falls within the *Scaphites hippocrepis* II ammonoid biozone, dated to 82.00 Ma [34].

Node 952 | SG Dipelta | *Dipelta-Kolkwitzia* | 33.71 Ma.
Fossil taxon and specimen. *Diplodipelta reniptera* [UM 33621: University of Michigan, Ann Arbor MI, USA] from the Florissant Formation, Mormon Cr, Ruby, CO, USA [78].
Phylogenetic justification. Based on the distribution of morphological characters in the phylogeny of Caprifoliaceae, this fossil is considered to be sister to the genus *Dipelta* which provide in turn the arguments for this assignment [78, 79].
Minimum age. 33.71 Ma.
Age justification. The Florissant Formation has been correlated with the early Chron C13r, based on its reversed polarity and a $^{40}Ar/^{39}Ar$ date of 34.07 Ma ± 0.10 Myr from the upper part of the section [80, 81]. The minimum age for the Florssant flora can be established on the C13r-C13n boundary, which has been dated to 33.71 Ma, based on the combined age model in Vandenberge et al. [50].

Node 983 | CG Araliaceae core | MRCA: *Cussonia-Tetralpasandra* | 37.3 Ma.
Fossil taxon and specimen. *Dendropanax eocenensis* [W1107': Indiana University paleobotanical collection] from the Claiborne Formation at Warman Clay Pit (2 miles west of Como), Tennessee, USA.
Phylogenetic justification. Phylogenetic relationships were discussed by Martinez-Millan [82].
Minimum age. 37.3 Ma.
Age justification. The Claiborne Formation is commonly attributed a middle Eocene age, but material evidence is rarely presented in support this. Taylor [83] provides a summary of the palynostratigraphic evidence, citing Elsik as supporting a middle Eocene age based on juglandaceous palynomorphs including *Plicatopollis, Platycarya, Platycaryapollenites, Carya* and *Casuarinidites*. Their presence may be indicative of a late Eocene age, though they occur in low frequency and, thus, Taylor [83] argues instead for a Middle Eocene age based on the presence of *Nuxpollenites terminalis* and *Amanoa* type palynomorphs in even the uppermost Claiborne Formation [84]. Thus, we establish a minimum age constraint on the age of *Solanites crassus* based on the Bartonian-Priabonian Boundary, dated to 37.8 Ma ± 0.5 Myr based on the combined age model of Vandenberghe et al. [50], thus, 37.3 Ma.

Node 1009 | SG Asteraceae minus Barnadesia | MRCA: *Barnadesia-Helianthus* | 41.5 Ma.
Fossil taxon and specimen. *Tubulifloridites antipodica* from onshore deposits taken from a paleochannel at Koingnaas, on the west coast of South Africa.
Phylogenetic justification. This fossil is accepted to belong to the Barnadesioides despite its actual placement in the subfamily is ambiguous but this is not a challenge to our analyses[82,85].
Minimum age. 41.5 Ma.
Age justification. These occurrences are, described to occur alongside the planktic forams *Globigerinatheka index* and *Turborotalia centralis* [85]. *Globigerinatheka index* is known to range from 42.9 - 34.3 Ma [86], but *Turborotalia centralis* is a junior synonym of *Turborotalia pomeroli*, which is known to range from 42.4-41.5 Ma [86]. Thus, the minimum age constraint on *Tubulifloridites antipodica* is 41.5 Ma. The newly described *Tubulifloridites lilliei* arguably slightly predates the other fossils of this palynological genus in its age[58].

Node 1047 | SG Aquifoliaceae | MRCA: *Ilex- Phyllonoma* | 61.6 Ma.
Fossil taxon and specimen. *Ilex hercynica* [MAI, Nr. 6004: Zentralsammlung Zentrales Geologisches Institut
Berlin] described from Walkmühle in Gonna, Germany [87-89].
Phylogenetic justification. Phylogenetic relationships were discussed by Martinez-Millan [82].
Minimum age. 61.6 Ma
Age justification. Knobloch et al. [90] assert an early Palaeocene age range for *Ilex hercynica*, affording a minimum constraint of 61.6 Ma based on the Danian-Salandian Boundary and the combined age model of Vandenberghe et al. [50].

Node 1075 | CG Solanales | MRCA: *Solanum-Montinia* | 37.3 Ma.
Fossil taxon and specimen. *Solanites crassus* [USNM-39949: Smithsonian Institution National Museum of Natural History, Washington DC, USA] from Holly Springs Sand in the Claiborne Formation at Mill Creek, at a railroad cut north of Shandy, Hardeman County, Tennessee, USA [91].

Phylogenetic justification. The phylogenetic interpretation of the fossil record of Solanales have been explored recently including the fossil incorporated here .
Minimum age. 37.3 Ma.
Age justification. The Claiborne Formation is commonly attributed a middle Eocene age, but material evidence is rarely presented in support this. Taylor [83] provides a summary of the palynostratigraphic evidence, citing Elsik [84] as supporting a middle Eocene age based on juglandaceous palynomorphs including *Plicatopollis, Platycarya, Platycaryapollenites, Carya* and *Casuarinidites*. Their presence may be indicative of a late Eocene age, though they occur in low frequency and, thus, Taylor [83] argues instead for a Middle Eocene age based on the presence of *Nuxpollenites terminalis* and *Amanoa* type palynomorphs in even the uppermost Claiborne Formation [84]. Thus, we establish a minimum age constraint on the age of *Solanites crassus* based on the Bartonian-Priabonian Boundary, dated to 37.8 Ma ± 0.5 Myr based on the combined age model of Vandenberghe et al. [50], thus, 37.3 Ma,

Node 1098 | CG Ericales core | MRCA: *Arbutus-Camellia* | 85.8 Ma.
Fossil taxon and specimen. *Paleoenkianthus sayrevillensis* [CUPC 1100: L. H. Bailey Hortorium, Cornell University, Ithaca NY, USA] from the South Amboy Fire Clay of the Raritan Formation, of which outcrops are exposed in the Old Crossman Clay Pit in Sayreville, New Jersey.
Phylogenetic justification. The phylogenetic relationships of this fossil have been tested based on morphological evidence [93].
Minimum age. 85.8 Ma
Age justification. Clarke et al. [1] argued that a minimum constraint on the age of this deposit could be established from Santonian-Campanian Boundary, however, Massoni et al. [53] argue that a tighter correlation can be established with better rocks attributable to the CC13-14 Nannofossil zones in South Carolina, indicating a minimum age of 86.3 Ma ± 0.5 Myr, thus, 85.8 Ma.

Node 1115 | CG Cornales | MRCA: *Cornus-Petalyonyx* | 85.8 Ma.
Fossil taxon and specimen. *Tylerianthus crossmanensis* [CUPC 1047: L. H. Bailey Hortorium, Cornell University, Ithaca NY, USA] from the South Amboy Fire Clay of the Raritan Formation, of which outcrops are exposed in the Old Crossman Clay Pit in Sayreville, New Jersey [94].
Phylogenetic justification. The phylogenetic relationships of this fossil have been clarified [95, 96].
Minimum age. 85.8 Ma
Age justification. Clarke et al. [1] argued that a minimum constraint on the age of this deposit could be established from Santonian-Campanian Boundary, however, Massoni et al. [53] argue that a tighter correlation can be established with better rocks attributable to the CC13-14 Nannofossil zones in South Carolina, indicating a minimum age of 86.3 Ma ± 0.5 Myr, thus, 85.8 Ma.

Node 1171 | CG Buxales | MRCA: *Didymeles-Buxus* | 100.1 Ma.
Fossil taxon and specimen. *Spanomera marylandensis* [PP42978: Field Museum, Chicago IL, USA] described by Drinnan et al. from the Potomac Formation at West Brothers clay pit, east of Washington, DC, Prince Georges County, eastern Maryland (late Albian), assigned to Palynozone IIB.
Phylogenetic justification. Based on similarities in the inflorescences and the striate tricolpate pollen, this fossil is likely related to the extant Buxaceae [98].
Minimum age. 100.1 Ma

Age justification. In the absence of further constraint on where within Zone IIB *Spanomera marylandensis* was recovered, we take the upper age constraint on the age of this zone to establish a minimum age constraint. Massoni et al. [53] argued that deposits lower within IIB were middle Albian, but Hochuli et al. [99] have demonstrated that at least some of Zone IIB is Upper Albian. Thus, we establish our minimum age constraint on the Albian-Cenomanian Boundary, dated to 100.5 Ma 0.4 Myr [34].

Node 1175 | CG Proteales | *Nelumbo-Platanus* | 107.59 Ma.
Fossil taxon and specimen. *Aquia brookensis* [PP4295: Field Museum, Chicago IL, USA] described by Crane et al. [100] from the Potomac Formation at Bank, near Brooke, Virginia, USA.
Phylogenetic justification. Doyle [98] recognized *Aquia* (combination of *Sapindopsis variabilis, Platanocarpus brookensis, Aquia brookensis*) as a stem group member of the genus *Platanus* and the widespread Albian leaf fossil *Nelumbites* as related to the genus *Nelumbo*.
Minimum age. 107.59 Ma
Age justification. Palynological correlations place Brooke in lower Subzone II-B, for which Massoni et al. [53] provide a very detailed biostratigraphic justification for a middle Albian age, to which we refer readers. Thus, a minimum constraint for Proteales may be established on the middle-late Albian Boundary, which coincides with the base of the *Diploceras cristatum* Biozone, dated to 107.59 Ma [34].

Node 1182 | SG Menispermaceae | MRCA: *Menispermum-Ranunculus* | 83.41 Ma.
Fossil taxon and specimen. *Prototinomiscium testudinarum* and *P. vangerowii* from Klikov Formation, Czech Republic [89].
Phylogenetic justification. The phylogenetic relationships of the fossil were clarified in a study testing them as calibration points for divergence time estimates of Menispermacae [101].
Minimum age. 83.41 Ma
Age justification. An Upper Turonian-Santonian age was established for the Klikov Formation by Pacltova [102] and Knobloch [103]. Thus, a minimum constraint can be established on the Santonian-Campanian Boundary, coincident with the base of the *Scaphites Leei III* Zone, dated to 83.64 Ma ± 0.23 Myr [34], thus, 83.41 Ma.

Node 1193 | CG Monocots | MRCA: *Acorus-Saccharum* | 112.6 Ma.
Fossil taxon. The earliest records of *Liliacidites* occur at the Trent's Reach Locality of the Potomac Group, attributable to the Albian Zone I [99].
Phylogenetic justification. Doyle et al. [104] identified pollen referred to the genus *Liliacidites* (but not Similipollis) as represesentative of the monocot stem, making it the oldest secure record of the monocot total group (see [98]).
Minimum age. 112.6 Ma.
Age justification. In the absence of further stratigraphic constraint, these earliest records of *Liliacidites* can be constrained in age by the Aptian-Albian Boundary, dated to 113.0 ma ± 0.4 Myr, thus, 112.6 Ma.
Discussion. Doyle et al. [104] highlight that, despite decades of sampling of the Hauterivian and Barremian of England, no clear representatives of *Liliacidites* pollen have been recovered [105], perhaps implying that the earliest records from the Albian are a close approximation of their antiquity. Because of the position of monocots in

our molecular tree we consider *Liliacidites* to be nested within monocots, and use it to calibrate the monocot crown node.

Node 1209 | SG Musaceae | MRCA: *Musa-Maranta* | 74.6 Ma.
Fossil taxon and specimen. *Spirematospermum chandlerae* has been described from isolated seeds and groups of seeds from the Neuse River locality, Black Creek Formation, southwest of Goldsboro, Wayne County, North Carolina, USA.
Phylogenetic justification. The phylogenetic relationships of this fossil have discussed in previous studies [106].
Minimum age. 74.6 Ma.
Age justification. Widely reported, but not evidenced as Late Cretaceous (Early Campanian) in age [107], the Black Creek Formation has been assigned to the *Exogyra ponderosa* Biozone which occurs beneath the *Didymoceras cheyennense* Tethyan ammonoid biozone [108], the base of which is dated to 74.6 Ma [34].

Node 1211 | CG Arecales | MRCA: *Elaeis-Chamaedorea* | 83.41 Ma.
Fossil taxon and specimen. *Sabalites carolinensis* [PAL 175717/P 38208: Smithsonian Museum of Natural History; Washington DC, USA] described from the Middendorf Arkose Member of Black Creek Formation near Langley, Aiken County, South Carolina [109].
Phylogenetic justification. The phylogenetic relationships of this fossil have discussed in Hertweck et al. [106] and Iles et al.
Minimum age. 83.41 Ma
Age justification. Berry's view that the Middendorf was merely a distinct facies within the Black Creek Formation, rather than a stratigraphically distinct unit, has been rejected. Sohl and Owens [111] subdivided the Upper Cretaceous of Carolina coastal plain into three lithostratigraphic units, the Middendorf, Black Creek and Peedee Formations, raised the Black Creek to group status and subdivided this into three unconformity-bound formations, viz. in stratigraphic sequence, the Tar Heel, Bladen and Donoho Creek formations. Evidently, *Sabalites carolinensis* was recovered from what is now recognized as the Middendorf Formation, and a minimum age constraint can be established on the boundary between the Middendorf and Tar Heel Formations. The Middendorf is commonly considered Santonian in age, however, little material evidence has been presented in support of this, in part a consequent of the complex history of stratigraphic divisions at outcrop, in subsurface and offshore [112]. Habib and Miller [113] established an age 'not younger than Campanian' on the basis of dinoflagellate biostratigraphy, but following the stratigraphic scheme outlined Campbell and Grohn [112], the Middendorf is older that the Shepherd Grove Formation and, therefore, following the stratigraphy of Christopher and Prowell [108], must be no younger than Santonian. Thus, we may established a minimum age constraint on the *Sabalites carolinesis* based on the Santonian-Campanian Boundary, coincident with the base of the *Scaphites leei III* Zone, dated to 83.64 Ma ± 0.23 Myr [34], thus, 83.41 Ma.

Node 1221 | CG Orchidaceae | MRCA: *Onicidium-Phaelaenopsis* | 17.82 Ma.
Fossil taxon and specimen. *Meliorchis caribea* [MCZ-31141: Museum of Comparative Zoology (Harvard University), Cambridge MA, USA], recovered from a mine east of Santiago, Cordillera Septentrional, Dominican Republic [114].
Phylogenetic justification. This amber fossil is composed of a pollinium attached to the wing of a stingless bee showing some similarities to the extand genus *Ligeophila*

[114]. Pollinium structure provide unequivocal evidence for the occurrence of Orchidaceae at this time.
Minimum age. 17.82 Ma.
Age justification. The age of Dominican Amber has been the subject of much speculation, but its dating has been best constrained on the basis of plantik forams, to late Early to early Middle Miocene [115] who list a series of biostratigraphically important species of planktic foraminifera, the stratigraphic ranges of which are not entirely compatible. However, the contemporaneous sediments contain including *Catapsydrax dissimillis*, the last appearance of which has been dated to 17.62 Ma [86], providing a minimum constraint on the age of *Meliorchis caribea*.

Node 1222 | CG Lillales | *Trillium-Lilium* | 18.7 Ma.
Fossil taxon and specimen. *Luzuriaga contortus* [L24916: Palynological Type Collection of th eNew Zealand Institute of Geological and Nuclear Sciences], based on leaf remains, from the Foulden Hill Diatomite, near Middlemarch, Otago, New Zealand [116].
Phylogenetic justification. The preserved structure shows high similarity to structures observed in extend this of the genus *Luzuriaga* and therefore the fossil is assigned unequivocally to the Alstromeriaceae.
Minimum age. 18.7 Ma.
Age justification. The deposit has been well studied and its maximum age is constrained by radiometric dating to 23.3 Ma ± 0.2 Ma [117]. The deposit has been attributed to the Waitakian, but a minimum age constraint must also encompass evidence of early Otaian elements of the flora and fauna, including *Assamiapollenites incognitus* which extends into the Otaian *Proteacidites isopogiformis* Zone [118]. The Otaian-Altonian Boundary has been dated to 18.7 Ma [119, 120].

Node 1224 | CG Dioscoreales-Pandanales| MRCA: *Dioscorea-Croomia* | 85.8 Ma.
Fossil taxon and specimen. *Mabelia connatifila* [CUPC 1255: L. H. Bailey Hortorium Paleobotanical Collection,
Cornell University, Ithaca NY, USA] from the South Amboy Fire Clay Member of the Raritan Formation at the Old Crossman clay pit in Sayreville, New Jersey, USA [121].
Phylogenetic justification. The phylogenetic assignment is based on the phylogenetic hypothesis reconstructed by Gandolfo et al. [121].
Minimum age. 85.8 Ma.
Age justification. Clarke et al. [1] argued that a minimum constraint on the age of this deposit could be established from Santonian-Campanian Boundary, however, Massoni et al. [53] argue that a tighter correlation can be established with better rocks attributable to the CC13-14 Nannofossil zones in South Carolina, indicating a minimum age of 86.3 Ma ± 0.5 Myr, thus, 85.8 Ma.

Node 1226 | CG Alismatales | MRCA: *Orontium-Hydrocharis* | 96.24 Ma.
Fossil taxon and specimen. *Mayoa portugallica* [S136663: Swedish Museum of Natural History Palaeobotanical Collection], from the Torres Vedras flora of the Figueira da Foz Formation [122], as the oldest record of crown-Alismatales.
Phylogenetic justification. Magallón et al. [13] argue that these striate and inaperturate pollen grains are similar to those of Monsteroideae (Araceae), such as *Holochlamys* and *Spathiphyllum*.
Minimum age. 96.24 Ma
Age justification. The Torres Vedras flora has been considered Late Barremian to Early Aptian in age [71, 123]. However, recent evidence suggests that they are

considerably younger, within the 'Upper Almargem' Formation overlying a late Aptian to early Albian unconformity [124]. Despite the uncertainty, an unequivocal minimum age is provided by the appearance of ostracod *Fossocytheridea merlensis* in the overlying Canecas Formation, attributable to the base of the Middle Cenomanian [125, 126], which coincides with the base of the Conlinoveras gilberti Zone, dated to 96.24 Ma [34].

Node 1228 | MRCA: *Alisma-Potamogeton* | 66 Ma.
Fossil taxon and specimen. *Cardstonia tolmanii* [UAPC-ALTA S55138: University of Alberta Paleobotanical
Collections] from the St. Mary River Formation [127].
Phylogenetic justification. Ridley and Stockey (2004) provided a convincing argument that this fossil shows closest similarity to extant genera of Alismatales. This interpretation is accepted here.
Minimum age. 66 Ma.
Age justification. Riley and Stockey [127] attribute *Cardstonia tolmanii* to the Late Campanian–Early Maastrichtian (Upper Cretaceous), though there is little direct evidence to support this. However, the St. Mary River Formation is chronostratigraphically equivalent to the Horseshoe Canyon Formation into which it intergrades and which in turn is constrained minimally by the C30n magnetozone [128], the top of which is slightly older than the Maastrichtian-Paleogene Boundary, dated to 66.0 Ma [34].

Node 1232 | CG Araceae | MRCA: *Orontium-Xanthosoma* | 76 Ma.
Fossil taxon and specimen. *Lysichiton (Araciphyllites) austriacus* [NHMW 1999B0057/0183: Natural History Museum, Vienna, Austria], from the Grünbach Formation of Austria, is the oldest record of this clade [129, 130]
Phylogenetic justification. This fossil was interpreted as closely related to extant species of the Orontioideae (family Araceae) by Bogner et al. [129]. This assignment is supported by the occurrence of several putative related fossil taxa at around the same time [131].
Minimum age. 76 Ma.
Age justification. The Grunbach Formation can be attributed to the *Globotruncata elevata* planktic foram biozone and the UC15 calcareous nannofossil biozone [130, 132]. The overlying Piesting Formation has been attributed to the UC16 nannofossil biozone [130, 132] and, thus, the age of the Grunbach Flora can be constrained by the UC15-UC16 boundary which is estimated to be 76 Ma [34].
Node 1237 | CG Laurales | MRCA: *Laurus-Calycanthus* | 107.59 Ma.
Fossil taxon and specimen. *Virginianthus calycanthoides* [PP43703: Field Museum, Chicago IL, USA] from the Albian of Puddledock, Virginia [133].
Phylogenetic justification. Doyle and colleagues [104, 134] identify *Virginianthus calycanthoides* from the Albian of Puddledock, Virginia [133], as the sister group to Calycanthaceae, or to the remaining Laurales, in either instance the oldest record of crown Laurales. Massoni et al. [53] also recognise *Cohongarootonia hispida* from the Puddledock Flora [135] as the oldest record of the total group comprised of Lauraceae, Monimiaceae, and Hernandiaceae, in other words the crown node of the clade of Laurales excluding Calycanthaceae [53]. We accept both of these as records of crown-Laurales, deeming evidence of their membership of more derived clades as insufficiently robust to be used as a basis for calibration.
Minimum age. 107.59 Ma.

Age justification. Massoni et al. [53] reason that the sediments in the Puddledock Locality are definitively early Albian based on the presence of reticulate tricolpate pollen and *Clavatipollenites rotundus* (aff. *Retimonocolpites dividuus* [136] but not striate tricolpates, which occur later in the early Albian.

Node 1248 | CG Magnoliales | MRCA: *Magnolia- Myristica* | 110.87 Ma.
Fossil taxon and specimen. *Endressinia brasiliana* [MB. PB. 2001/1455: Museum of Natural History, Institute of Paleontology, Berlin, Germany], from the Crato Formation of Brazil [137].
Phylogenetic justification. Massoni et al. identify both *Schenkeriphyllum glanduliferum* and *Endressinia brasiliana*, both from the Crato Formation of Brazil [137, 138], as the oldest records of crown Magnoliineae, the sister clade of Myristicaceae, based on the phylogenetic analyses [98, 134, 138].
Minimum age. 110.87 Ma.
Age justification. Clarke et al. [1] argued that the age of the Crato Formation could not be constrained to being definitively older than Albian based on pollen [140], ostracod [141], and dinoflagellate [142] biostratigraphy and, in the absence of further evidence, established a minimum constraint on the Albian-Cenomanian boundary. Massoni et al. [53] argued for an Aptian age for the Crato Formation based on evidence from Heimhofer and Hochuli [142] but, unfortunately, these authors do not present evidence that can discriminate against a possible early Albian age for the Crato Formation, as acknowledged by Mohr et al. [138]. While the evidence suggests at worst, an early Albian age for the Crato Formation, and so it is possible to derive a minimum age interpretation for the Formation based on the Early-Middle Albian Boundary, which coincides approximately with the base of the *Douvilleiceras mammillatum* ammonite biozone, dated to 110.87 Ma [34].
Discussion. *Archaeanthus linnenbergii* was recognized as a further putative stem group Magnoliaceae but it is younger than *Endressinia* [53, 98].

Node 1256 | SG Canellales | MRCA: *Piper-Canella* | 125.9 Ma.
Fossil taxon and specimen. *Walkeripollis gabonensis* [Single-pollen grain preparation 2963-27: University of California (Berkeley) Museum of Paleontology], Cocobeach sequence near N'Toum, Gabon [53].
Phylogenetic justification. Massoni et al. [53] identify *Walkeripollis gabonensis* as the oldest record of Canellales, based on phylogenetic of pollen grain characters that resolve this pollen species as stem-Winteraceae [134], one of the two families that comprise Canellales, based on the presence of permanent tetrads, rounded aperture shape, as well as characteristic pore sculpture.
Minimum age. 125.9 Ma.
Age justification. Massoni et al. [53] establish a pre-Aptian age for *Walkeripollis gabonensis* based on the correlation of its occurrence in the Elf-Aquitaine palynological Subzone C-VIIc of the Cocobeach sequence near N'Toum, Gabon, to better-dated late Barremian sequences elsewhere. Thus, a minimum age constraint is established on the Barremian-Aptian boundary that has been dated to 126.3 Ma ±0.4 Myr, thus, 125.9 Ma.
Discussion. Massoni et al. [53] proposed records of more derived clades within Canellales (e.g. *Appomattoxia*, including the pollen of *Tucanopollis* and *Transitoripollis* types, from the Puddledock, Virginia [143], and Torres Vedras, Portugal [71, 144]). However, these are not sufficiently convincing to use as a basis for calibration. For example, *Appomattoxia* maybe more closely related to *Ceratophyllum* and Chloranthaceae [98].

Node 1260 | SG Saururus | MRCA: *Saururus-Houttuynia* | 44.3 Ma.
Fossil taxon and specimen. *Saururus tuckerae* [UAPC P1631 Bbot a: University of Alberta (Edmonton) Paleobotanical Collections] from the Middle Eocene Princeton Chert, British Columbia, Canada.
Phylogenetic justification. Massoni et al. [53] follow Smith and Stockey [145] in identifying *Saururus tuckerae* as the oldest record of total group *Saururus*. Based on tens of flowers and a partial inflorescence, the flower structure and pollen are characteristic of Saururaceae (Piperales), and phylogenetic analyses resolved *S. tuckerae* as the sister clade to extant *Saururus* [145].
Minimum age. 44.3 Ma.
Age justification. The Princeton Chert is part of the Allenby Formation which has been the subject of a number of absolute dating studies yielding age estimates of 48 Ma ± 2 Myr [146, 147], between 47 Ma ± 2 Myr and 50 Ma ± 2 Myr [148], between 46.2 Ma ± 1.9 Myr and 49.4 Ma ± 2 Myr [149], and 52.08 Ma ± 0.12 Myr [150] for the Allenby Formation. We follow Massoni et al. [53] in basing our minimum constraint based on the youngest age Interpretation of the youngest radiometric age estimate, viz. 44.3 Ma

Node 1268 | CG Chloranthales | MRCA: *Chloranthus-Hedyosmum* | 92.8 Ma.
Fossil taxon and specimen. *Pennipolis* plant [151] based on material from Vale de Agua, Buarcos, Portugal [152].
Phylogenetic justification. *Pennipolis* was found as sister to the crown group Chloranthaceae [153], whereas the Aptian to Cernomanina *Asteropollis* mesofossils may represent both crown and stem group Chloranthaceae [98].
Minimum age. 92.8 Ma.
Age justification. Clarke et al. [1] established a minimum age for Vale de Agua, which is assigned to the Figueira da Foz Formation [154-156] based on the overlying Costa d'Arnes Formation, the oldest ammonites in which include *Calycoceras naviculare* [157], indicative of the naviculare biozone, the dating of which in error of the top of the Cenomanian, dated 93.6 Ma ± 0.8 Myr [34], thus, 92.8 Ma.
Discussion. Other early Cretaceous fossils with possible relationships to Chloranthaceae, such as *Zlatkcarpus* and *Canrightia* are younger than *Pennipolis* [98].

Node 1272 | SG Schisandraceae | MRCA: Trimenia-Kadsura | 107.59 Ma.
Fossil taxon and specimen. *Anacostia virginiensis* based on material from Kenilworth, Maryland, Puddledock, Virginia.
Phylogenetic justification. The assignment is based on the result of maximum parsimony based reconstruction of the phylogenetic relationships of this fossil to extant angiosperms that recovered this taxon as nested between *Trimenia* and the clade comprising *Illicium* and *Schisandra* [153].
Minimum age. 107.59 Ma.
Age justification. Massoni et al. [53] reason that the sediments in the Puddledock Locality are definitively early Albian based on the presence of reticulate tricolpate pollen and *Clavatipollenites rotundus* (aff. *Retimonocolpites dividuus* [136]) but not striate tricolpates, which occur later in the early Albian. Therefore, they constrain minimally the age of the *A. virginiensis* by the Middle-late Albian boundary, which coincides with the base of the *Diploceras cristatum* biozone which has been dated to 107.59 Ma [34].

Discussion. *Anacostia*, reportedly from the early and middle Albian of Buarcos, Famalicão, and Vale de Agua (Portugal), Puddledock (Virginia, USA), and Kenilworth (Maryland, USA) was recognized as the oldest fossil record of the Austrobaileyales [98, 153]. Doyle and Endress [153] *identified Anacostia portugallica* and *A. teixeirae* as early Albian and, therefore the oldest species belonging to this lineage. However, the minimum age interpretation of these localities the Figueira da Foz Formation cannot be constrained minimally to more than 92.8 Ma (see above). However, the minimum age constraint on *A. virginiensis* from the Puddledock Locality is older.

Node 1273 | SG *Illicium* | MRCA: *Illicium-Schisandra* | 85.44 Ma.
Fossil taxon and specimen. *Illiciospermum pusillum* [1700b-127: Komarov Botanical Institute, St. Petersburg, Russia.], known from seeds from the Cenomanian-Turonian of Kazahkstan [158].
Phylogenetic justification. *Illiciospermum pusillum,* known from seeds that preserve a structure resembling the strophiole of *Illicium* [158].
Minimum age. 85.44 Ma.
Age justification. There is a paucity of evidence supporting the age interpretation of the Sarbay Quarry near Rudnyy, Kustanay Region, north-western Kazakhstan. Frumin and Friis (1996) describe the sediments, including the plant-bearing bed, as belonging to the Shet-Irgiz Formation of Cenomanian-Turonian age [159], overlain by marine sands of the Santonian-Campanian Ayat Suite containing *Inoceramus cardissoides* [159]. The *Inoceramus cardissoides* Zone falls fully within the *Sigalia carpatica* Planktic foraminiferal Zone [160], the range end of which is dated to 85.44 Ma [34].

Node 1276 | SG Cabombaceae | MRCA: *Nymphaea-Cabomba* | 110.87 Ma.
Fossil taxon and specimen. *Pluricarpellatia peltata* [MB.Pb. 2000/80: Museum of Natural History, Berlin, Germany], from the Crato Formation of Brazil [161]
Phylogenetic justification. *Pluricarpellatia peltata* has been considered phylogenetically and resolved as members of the lineage leading to *Cabomba* after it diverged from *Nymphaea* [162].
Minimum age. 110.87 Ma.
Age justification. Clarke et al. [1] argued that the age of the Crato Formation could not be constrained to being definitively older than Albian based on pollen [140], ostracod [141], and dinoflagellate [142] biostratigraphy and, in the absence of further evidence, established a minimum constraint on the Albian-Cenomanian boundary. Massoni et al. [53] argued for an Aptian age for the Crato Formation based on evidence from Heimhofer and Hochuli [142] but, unfortunately, these authors do not present evidence that can discriminate against a possible early Albian age for the Crato Formation, as acknowledged by Mohr et al. [138]. While the evidence suggests, at worst, an early Albian age for the Crato Formation, it is possible to derive a minimum age interpretation for the Formation based on the Early-Middle Albian Boundary, which coincides approximately with the base of the *Douvilleiceras mammillatum* ammonite biozone, dated to 110.87 Ma [34].
Discussion. Magallon et al. [13] derive a minimum constraint from *Monetianthus mirus* which they recognize as a representative of the Nymphaeaceae stem lineage and, thus, use it as the basis of a minimum constraint on the age of total-group Nymphaceaceae at 125 Ma. However, Clarke et al. [1] demonstrated that the minimum age of the host deposit, Vale de Água, Portugal [163, 164] is 93.9 Ma [34]. However, there are other, potentially older records of Nymphaeaceae and, more specifically, the crown clade circumscribed by *Nymphaea-Cabomba*. Clarke et al. [1]

identified much older, but more equivocal records, as well as the oldest unequivocal records, viz. *Pluricarpellatia peltata* from the Crato Formation of Brazil [161] and *Scutifolium jordanicum* from the Jarash Formation (Kurnub Group) of Jordan [162], both of which have been considered phylogenetically and resolved as members of the lineage leading to *Cabomba* after it diverged from *Nymphaea* [162]. *Scutifolium jordanicum* was used to establish a minimum age for crown-Nymphaeales at 105 Ma by Smith et al. [165], and for total-group Cabombaceae at 105 Ma by Zanne et al. [166]. The Jarash Formation can be dated minimally to 95 Ma (96.1 Ma ± 1.1 Myr in [167], but the Crato Formation is older.

Node 1278 | CG Nymphaeaceae | MRCA: *Nymphaea - Nuphar* crown | 92.8 Ma.
Fossil taxon and specimen. *Monetianthus mirus* [S122015: Palaeobotanical Collectons, Swedish Museum of Natural History, Stockholm, Sweden] from Vale de Agua, Portugal [164].
Phylogenetic justification. Doyle and Endress [153] identify *Monetianthus mirus*, a coalified flower from Vale de Agua, Portugal, as a stem member of the clade Barclaya-Nymphaeoideae, to the exclusion of *Nuphar*.
Minimum age. 92.8 Ma.
Age justification. Clarke et al. [1] established a minimum age for Vale de Agua, which is assigned to the Figueira da Foz Formation [154-156], based on the overlying Costa d'Arnes Formation, the oldest ammonites in which include *Calycoceras naviculare* [157], indicative of the *naviculare* Biozone, the dating of which in error of the top of the Cenomanian, dated 93.6 Ma ± 0.8 Myr [34], thus, 92.8 Ma.

Node 1279 | CG Acrogymnosperms | MRCA: *Ginkgo-Pinus* | 308.14 Ma – 365.629 Ma.
Fossil taxon and specimen. *Cordaites iowensis* [UM4616: University of Michigan and Illinois Geological Survey, Ann Arbor MI, USA] from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa, USA [22].
Phylogenetic justification. Clarke *et al*. [1] identify cordaitean coniferophytes as the oldest records of the *Ginkgo-Pinus* clade, the oldest whole plant reconstruction of which is *Cordaites iowensis* from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa [22].
Minimum age. 308.14 Ma.
Maximum age: 365.629 Ma.
Age justification. Janousek and Pope [23] argue that the Laddsdale Coal is equivalent to the Bluejacket Coal of Oklahoma, which occurs as part of the Bluejacket Sandstone Member, underlying the Inola Limestone, part of the Inola Cyclothem of the Krebs subgroup of the Cherokee Group, characterized by the occurrence of the conodonts *Idiognathodus amplificus, Idiognathodus podolskensis and Neognathodus asymmetricus* [24]. The Inola cyclothem falls fully within the *Idiognathodus amplificus/ Idiognathodus obliquus* biozone [25]. This is indicative of the *Neognathodus medexultimus-Streptognathodus concinnus* (Pc10) biozone, certainly older than the *Neognathodus roundyi – Streptognathodus cancellosus* (Pc11) biozone [25, 26]. The base of Pc10 is bracketed by an older age constraint of 312.01 Ma ± 0.37 Myr and the base of Pc11 is bracketed by a younger age constraint of 308.5 Ma ± 0.36 Myr in the Composite Standard of Davydov et al. [26], yielding a minimum age constraint of 308.14 Ma. A maximum is based upon the first appearance of seeds in the form of preovules which are attributable to the spermatophyte stem, the oldest interpretation of which is 365.629 Ma (see Spermatophyta).

Discussion. Zanne et al. [166] derive a minimum constraint from *Emporia lockardii* at 290.0 Ma which they recognize as a member of crown-Acrogymnospermae within a phylogenetic concept of the group in which, as here, cycads and *Ginkgo* comprise a clade.

Node 1280 | CG Conifers | MRCA: Pinus-Metasequoia | 147 Ma - 312.38 Ma.
Fossil taxon and specimen. *Araucaria mirabilis* [NHM V. 30953: Natural History Museum, London, UK], represented by cones, from Cerro Cuadrado petrified forest, La Matilde Formation, Patagonia, Argentina [168-171].
Phylogenetic justification. These fossils possess a 'vascular plexus' at the ovule base, ovuliferous scale vascularization, two vascular strands to the conescale complex and an embryo with two cotyledons, all characters established to distinguish *Araucaria* section *Bunya* of the Araucariaceae [170, 172], to which only extant *Araucaria bidwillii* belongs.
Minimum age. 147 Ma.
Maximum age: 312.38 Ma.
Age justification. The age of La Matilde Formation is poorly constrained as the stratigraphy is complex, although the volcanic deposits do allow radiometric dating. La Matilde Formation is overlain by volcanics dated to 157 Ma ± 10 Myr [173], and thus the minimum constraint on the divergence of crown Cupressophyta, total group Cupressophyta and crown Coniferae is 147 Ma. A maximum constraint can be based on *Cordaites iowensis*, a cordaitean coniferophyte from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa [22], is the oldest whole plant reconstruction for Coniferae. Janousek and Pope [23] argue that the Laddsdale Coal is equivalent to the Bluejacket Coal of Oklahoma, which occurs as part of the Bluejacket Sandstone Member, underlying the Inola Limestone, part of the Inola Cyclothem of the Krebs subgroup of the Cherokee Group, characterized by the occurrence of the conodonts *Idiognathodus amplificus, Idiognathodus podolskensis and Neognathodus asymmetricus* [24]. The Inola cyclothem falls fully within the *Idiognathodus amplificus/ Idiognathodus obliquus* biozone [25]. This is indicative of the *Neognathodus medexultimus-Streptognathodus concinnus* (Pc10) biozone, certainly older than the *Neognathodus roundyi – Streptognathodus cancellosus* (Pc11) biozone [25, 26]. The base of Pc10 is bracketed by an older age constraint of 312.01 Ma ± 0.37 Myr and the base of Pc11 is bracketed by a younger age constraint of 308.5 Ma ± 0.36 Myr in the Composite Standard of Davydov et al. [26], yielding a maximum of 312.38 Ma.
Discussion. This is the fundamental divergence of Coniferae into Cupressophyta, Gnetales and Pinaceae. The oldest secure records of the gnepine total group occur within the Yixian Formation of Liaoning, China, the minimum age of which is 121.8 Ma (see [1]). The oldest possible records of Cupressophyta total group include Triassic *Rissikia media* (Townrow, 1967) but it lacks the Podocarpaceae diagnostic feature of one ovule per cone scale, instead possessing two [1]. Other Triassic-Jurassic records are equally problematic [174-176].

Node 1282 | CG Gnetales | MRCA: *Gnetum-Welwitschia* | 119.6 Ma – 312.38 Ma.
Fossil taxon and specimen. *Eoantha zherikhinii* [Repository of the Institute of Biology and Pedology, Vladivostok, Russia], from the Zaza Formation at the Baisa locality in the upper reaches of the Vitim River in Lake Baikal [177].
Phylogenetic justification. This fossil is assigned as a representative of the Gnetales because of the presence of an ovule with a an exending micropylar tube similar to extend Gnetales and the presence of polyplicate pollen of *Ephedritopes-*

type (see [178]). *Eoantha* is one out of several Early Cretaceous fossils unequivocally assigned to the crown Gnetales [179]
Minimum age. 119.6 Ma.
Maximum age. 312.38 Ma.
Age justification. The Zaza Formation can be correlated with the Turga Formation, also of Transbaikalia based principally on common elements of their floral assemblages, including *Asteropollis asteroids*, *Dicotylophyllum pusilum*, *Baisa hirsuita, Podozamites, Schizolepis, Pseudolarix, Phoenicopsis, Czekanowskia rigida* and *Sphenobaiera* [177, 180-182]. The age of the Turga flora and Formation is based on the chronological distribution of *Asteropollis* type pollen, but correlation with the Yixian Formation of China is also supported strongly [180], allowing for refinement of the *Asteropollis*-derived ages. Correlation between Turga and Yixian is based on similarities in the floral assemblages of these two formations, with the shared presence of the species *Baisa hirsuita, Botrychites reheensis, Neozamites verchojanensis, Pityolepis pseudotsugaoides, Brachyphyllum longispicum, Scarbugia hilii, Ephedrites chenii, Carpolithus multiseminalis, Carpolithus pachythelis, Schizolepis, Baiera, Coniopteris, Ginkoites, Pityocladus*, *Pityospermum* and *Elatocladus* [177, 180, 183]. The shared presence of *Asteropollis asteroides* in Turga and Zaza can be used to constrain their age. The last appearance of *Asteropollis* pollen is in Antarctica [184] and is dated to the end-Campanian, at the latest 72.1 Ma ± 0.2 [34]. This minimum may be constrained further based on the correlation of the Zaza Formation through the Turga Formation to the Yixian Formation. The main fossil bearing beds in the Yixian Formation have been recently dated and may be as old as 129.2 Ma [41], however, in the absence of knowledge of the position of the fossils within the stratigraphy, relative to the sources of the absolute dates, a minimum age constraint can be derived from the Jiufontang Formation which overlies it. $^{40}Ar/^{39}Ar$ dating of a number of samples from the Jiufontang Formation has yielded an age of 120.3 ± 0.7 Ma for the volcanic tuffs [42], establishing a minimum constraint of 119.6 Ma for the age of the Yixian, Formation and, thus ultimately the Zaza Formation.

A maximum constraint can be based on *Cordaites iowensis*, a cordaitean coniferophyte from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa [22], is the oldest whole plant reconstruction for Coniferae. Janousek and Pope [23] argue that the Laddsdale Coal is equivalent to the Bluejacket Coal of Oklahoma, which occurs as part of the Bluejacket Sandstone Member, underlying the Inola Limestone, part of the Inola Cyclothem of the Krebs subgroup of the Cherokee Group, characterized by the occurrence of the conodonts *Idiognathodus amplificus, Idiognathodus podolskensis and Neognathodus asymmetricus* [24]. The Inola cyclothem falls fully within the *Idiognathodus amplificus/ Idiognathodus obliquus* biozone [25]. This is indicative of the *Neognathodus medexultimus-Streptognathodus concinnus* (Pc10) biozone, certainly older than the *Neognathodus roundyi – Streptognathodus cancellosus* (Pc11) biozone [25, 26]. The base of Pc10 is bracketed by an older age constraint of 312.01 Ma ± 0.37 Myr and the base of Pc11 is bracketed by a younger age constraint of 308.5 Ma ± 0.36 Myr in the Composite Standard of Davydov et al. [26], yielding a maximum of 312.38 Ma.

Node 1284 | MRCA: *Ginkgo-Cycas* | 264.7 Ma – 365.629 Ma.
Fossil taxon and specimen. *Crossozamia chinensis* [GP0027: Beijing Graduate School, China Institute of
Mining, Beijing, China], Lower Shihhotse Formation at Simugedong, Dongshan (East Hills), Taiyuan, north China [185].

Phylogenetic justification. Nagalingum et al. [186] identify *Crossozamia* as the oldest record of the *Cycas* lineage, based on megasporophylls that exhibit similarity to extant *Cycas* [185]. They argue against the interpretation of *Crossozamia* as the sister lineage of *Cycas* based on the presence of an estipulate leaf base and a terminal pinna found in the seedlings [187], instead favouring its assignment to the cycad stem. The arguments presented clearly raise doubts about the assignment of *Crossozamia* to crown-cycads, however, they do not provide definitive evidence of its exclusion from this clade and so *Crossozamia* may more appropriately be assigned to the cycad total group (i.e. we cannot discriminate between a stem or crown-cycad affinity based on the available evidence). In either instance, *Crossozamia* is the oldest record of the minimal clade comprised of *Gingko* and *Cycas*.
Minimum age. 264.7 Ma.
Maximum age. 365.629 Ma.
Age justification. The Lower Shihhotse Formation at Simugedong, Dongshan (East Hills), Taiyuan, north China [185] has been established biostratigraphically as Roadian-Wordian (middle Permian) [188] and, thus a minimum age constraint can be established on the Wordian-Capitanian Boundary which has been dated to 265.1 Ma ± 0.4 Myr [189]. Thus, the minimum age constraint on the *Cycas-Ginkgo* clade is 264.7 Ma. A maximum is based upon the first appearance of seeds in the form of preovules which are attributable to the spermatophyte stem, the oldest interpretation of which is 365.629 Ma (see Spermatophyta).

Node 1286 | CG Monilophytes | MRCA: *Psilotum-Adiantum* | 384.71 Ma – 451 Ma.
Fossil taxon and specimen. *Ibyka amphikoma* was recovered from the Manorkill Shale Member at Schoharie Creek directly below the spillway of Gilboa dam, Gilboa, Schoharie County, New York, Gilboa [14].
Phylogenetic justification. *Ibyka amphikoma* [14] is the oldest record of the equisetopsid lineage based on the phylogenetic analyses undertaken by Kenrick and Crane [3].
Minimum age. 384.71 Ma.
Maximum age. 451 Ma.
Age justification. *Ibyka amphikoma* was recovered from the Manorkill Shale Member, which is a lateral equivalent of the Windom Member, within the Moscow Formation of New York [18, 19], which falls fully within the *ansatus* conodont Biozone [20, 21] the top of which is dated to 385.41 Ma ± 0.7 Myr, thus, yielding a minimum age constraint of 384.71 Ma. The maximum constraint, following Clarke et al. [1], is based on the oldest occurrences of trilete spores, known from the Qusaiba-1 core from the Quasim Formation of northern Saudi Arabia [8]. The very oldest records preceed the earliest occurrences of the *A. barbata*, opening the possibility that they occur in the preceding *Tanuchitina fistulosa* biozone, though *T. fistulosa* does not occur. The oldest stratigraphic records within the core co-occur with the chitinozoan *Armoricochitina nigerica*, known to extend into the Caradoc, to within the biozone characterized by *Fungochitina spinifera* (= *Fungochitina fungiformis*) [9]. The base of the *F. spinifera* Zone falls within the *Dicranograptus clingani* Biozone (*Dicellograptus morrisi* Subzone) [10], the base of which is estimated at 451 Ma [11]. Thus, we establish our maximum for Tracheophyta at 451 Ma.
Discussion. Magallón et al. [13] established a minimum age constraint based on *Ibyka amphikoma* using the Givetian-Frasnian boundary, for which they provided a date of 385 Ma, though this has since been revised to 382.7 Ma ± 1 Myr [17]. However, we provide a more detailed stratigraphic justification for the age of *I. amphikoma* which allows for an older minimum age constraint.

Node 1287 | CG Lycophytes | MRCA: *Huperzia-Selaginella* | 392.1 Ma – 451 Ma.
Fossil taxon and specimen. *Leclercquia complexa* [CW092 (07 – 061): Collections of the Centre for Palynological Studies, Department of Animal and Plant Sciences, University of Sheffield, UK], from Campbellton Formation outcropping on the south shore of the Restigouche River, between Dalhousie and Campbellton, New Brunswick, eastern Canada [190].
Phylogenetic justification. Kenrick and Crane [3] identified *Leclercquia complexa* as the oldest member of Isoetopsida and crown Lycopodiophyta. This interpretation is supported by spore characteristics analysed phylogenetically by Wellman et al. [190].
Minimum age. 392.1 Ma.
Maximum age. 451 Ma.
Age justification. A Late Emsian age is often cited for the New Brunswick occurrences of identified *Leclercquia complexa* and, indeed, the *Stockmensella-Leclerqia* macroplant Biozone spans all but the earliest Emsian [17]. However, Wellman et al. [190] attribute their own material of *Leclercquia complexa* to the middle of the *Emphanisporites annulatus – Camarozonotriletes sextantii* Spore Assemblage Biozone which falls within the early part of the Emsian. In either instance, the earliest records of *Leclercquia complexa* fall fully within the Emsian, the end of which is dated to 393.3 Ma ± 1.2 Myr [17], yielding a minimum constraint of 392.1 Ma. The maximum constraint, following Clarke et al. (2011), is based on the oldest occurrences of trilete spores, known from the Qusaiba-1 core from the Quasim Formation of northern Saudi Arabia [8]. The very oldest records precede the earliest occurrences of the *A. barbata*, opening the possibility that they occur in the preceding *Tanuchitina fistulosa* biozone, though *T. fistulosa* does not occur. The oldest stratigraphic records within the core co-occur with the chitinozoan *Armoricochitina nigerica*, known to extend into the Caradoc, to within the Biozone characterized by *Fungochitina spinifera* (= *Fungochitina fungiformis*) [9]. The base of the *F. spinifera* Zone falls within the *Dicranograptus clingani* Biozone (*Dicellograptus morrisi* Subzone) [10], the base of which is estimated at 451 Ma [11]. Thus, we establish our maximum for Tracheophyta at 451 Ma.
Discussion. Magallon et al. [13] cite a minimum age of 385 Ma, based on the Middle-Upper Devonian Boundary, but our more detailed stratigraphy allows for an older minimum age interpretation of *Leclercquia complexa.*

Fossil calibration references

1. **Clarke J, Donoghue PCJ, Warnock RCM. 2011.** Establishing a timescale for plant evolution. *New Phytologist* **192**(1), 266-301.
2. **Kotyk ME, Basinger JF, Gensel PG, de Freitas TA. 2002.** Morphologically complex plant macrofossils from the Late Silurian of Arctic Canada. *American Journal of Botany* **89**(6), 1004-1013.
3. **Kenrick P, Crane PR. 1997.** *The origin and early diversification of land plants: a cladistic study*. Washington D.C., Smithsonian Institution Press.
4. **Klapper G, Murphy MA. 1974.** Silurian - lower Devonian conodont sequence in the Roberts Mountains Formation of central Nevada. *University of California Publications in Geological Sciences* **111**, 1-62.
5. **Thorsteinsson R. 1980.** Stratigraphy and conodonts of Upper Silurian and Lower Devonian rocks in the environs of the Boothia Uplift Canadian Arctic

Archipelago. 1. Contributions to stratigraphy. *Geological Survey of Canada Bulletin* (292), VIII-38.
6. **Uyeno TT. 1990.** Biostratigraphy and conodont faunas of Upper Ordovician through Middle Devonian rocks, eastern Arctic Archipelago. *Geological Survey of Canada, Bulletin* **401**, 1-211.
7. **Mayr U, Brent TA, de Freitas T, Frisch T, Nowlan GS, Okulitch AV. 2004**. Geology of Eastern Prince of Wales Island and Adjacent Smaller Islands, Nunavut. *Geological Survey of Canada, Bulletin* **574**, 1-88.
8. **Steemans P, Le Herisse A, Melvin J, Miller MA, Paris F, Verniers J, Wellman CH. 2009.** Origin and Radiation of the Earliest Vascular Land Plants. *Science* **324**(5925), 353-353.
9. **Paris F, Le Herisse A, Monod O, Kozlu H, Ghienne J-F, Dean WT, Vecoli M, Gunay Y. 2007**. Ordovician chitinozoans and acritarchs from southern and southeastern Turkey. *Revue de micropaléontologie* **50**(1), 81-107.
10. **Vandenbroucke TRA, Williams M, Zalasiewicz JA, Davies JR, Waters RA. 2008.** Integrated Upper Ordovician graptolite-chitinozoan biostratigraphy of the Cardigan and Whitland areas, southwest Wales. *Geological Magazine* **145**, 199-214.
11. **Cooper RA, Sadler PM. 2012.** The Ordovician Period. In *Geologic timescale 2012* (eds. Gradstein FM, Ogg JG, Schmitz M, Ogg G), pp. 489-523, Elsevier.
12. **Bonamo PM. 1977**. *Rellimia Thomsonii* (Progymnospermopsida) from Middle Devonian of New York State. *American Journal of Botany* **64**(10), 1272-1285.
13. **Magallón S, Hilu KW, Quandt D. 2013.** Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am J Bot* **100**(3), 556-573.
14. **Skog JE, Banks HP. 1973.** *Ibyka amphikoma*, gen et sp-n - new protoarticulate precursor from late Middle Devonian of New York State. *American Journal of Botany* **60**(4), 366-380.
15. **Bartholomew AJ, Brett CE. 2007.** Correlation of Middle Devonian Hamilton Group-equivalent strata in east-central North America: implications for eustasy, tectonics and faunal provinciality. *Geological Society, London, Special Publications* **278**(1), 105-131.
16. **Johnson JG, Klapper G, Sandberg CA 1985.** Devonian eustatic fluctuations in Euramerica. *Geological Society of America Bulletin* **96**(5), 567.
17. **Becker RT, Gradstein FM, Hammer O. 2012.** The Devonian Period. In *The geological timescale 2012* (eds. Gradstein F.M., Ogg J.G., Schmitz M., Ogg G.), pp. 559-601, Elsevier.
18. **Fisher DW., Isachsen YW., Rickard LV, Broughton JG, Offield TW. 1962.** *Geologic map of New York*. Albany, New York State Museum Sciece Service, Geological Survey.
19. **Rickard LV. 1964.** *Correlation of the Devonian rocks in New York State. Map and Chart Series 4*. Albany, New York State Musuem Science Service Geological Survey.
20. **Klapper G. 1981.** Review of New York Devonian conodont biostratigraphy. In *Devonian biostratigraphy of New York, Part I* (eds. Oliver W.A., Jr., Klapper G.), pp. 57-68, IUGS SDS.
21. **Kirchgasser WT. 2000.** Correlation of stage boundaries in the Appalachian Devonian, Eastern United States. *Courier Forschungsinstitut Senckenberg* **225**, 271-284.

22. **Trivett ML 1992.** Growt, architecture, structure, and relationships of *Cordaixylon iowensis* nov comb (Cordaitales). *International Journal of Plant Sciences* **153**(2), 273-287.
23. **Janousek TJ, Pope JP. 2014.** Petrology, petrography and conodont biostratigraphy of the Laddsdale Coal interval, along Whitebreast Creek, Bauer, Iowa. *GSA North-Central Section, 48th Annual Meeting, Abstracts* **16-5**.
24. **Heckel PH. 2013.** Pennsylvanian stratigraphy of Northern Midcontinent Shelf and biostratigraphic correlation of cyclothems. *Stratigraphy* **10**, 3-39.
25. **Barrick JE, Lambert LL, Heckel PH, Rosscoe SJ, Boardman DR. 2013.** Midcontinent Pennsylvanian conodont zonation. *Stratigraphy* **10**, 55-72.
26. **Davydov VI, Korn D, Schmitz MD. 2012.** The Carboniferous Period. In *The geologic time scale 2012* (eds. Gradstein F.M., Ogg J.G., Schmitz M., Ogg G.), pp. 603-651, Elsevier.
27. **Prestianni C. 2005.** Early diversification of seeds and seed-like structures. *Carnets De Geologie*, 33-38.
28. **Rothwell GW, Scheckler SE, Gillespie WH. 1989.** *Elkinsia* gen nov, a late Devonian gymnospermn with cupulate ovules. *Botanical Gazette* **150**(2), 170-189.
29. **Streel M, Scheckler SE. 1990.** Miospore lateral distribution in upper Fammenian alluvial lagoonal to tidal facies from eastern United States and Belgium. *Review of Palaeobotany and Palynology* **64**(1-4), 315-324.
30. **Streel M, Higgs K, Loboziak S, Riegel W, Steemans P. 1987.** Spore stratigraphy and correlation with faunas and floras in the type marine Devonian of the Ardenne-Rhenish regions. *Review of Palaeobotany and Palynology* **50**(3), 211-229.
31. **House MR., Gradstein FM. 2004.** The Devonian Period. In *A geologic timescale 2004* (eds. Gradstein FM, Ogg JG, Smith AG), pp. 202-221. Cambridge, Cambridge University Press.
32. **Hughes NF, McDougall AB. 1990.** Barremian-Aptian angiospermid pollen records from southern England. *Review of Palaeobotany and Palynology* **65**(1-4), 145-151.
33. **Judd WS, Olmstead RG. 2004.** A survey of tricolpate (eudicot) phylogenetic relationships. *American Journal of Botany* **91**(10), 1627-1644.
34. **Ogg JG., Hinnov LA. 2012.** Cretaceous. In *The geologic time scale 2012* (eds. Gradstein FM, Ogg JG, Schmitz M, Ogg G), pp. 793-853, Elsevier.
35. **Ogg JG. 2012.** Triassic. In *The geologic time scale 2012* (eds. Gradstein FM, Ogg JG, Schmitz M, Ogg G), pp. 681-730, Elsevier.
36. **Liu ZJ, Wang X. 2016.** A perfect flower from the Jurassic of China. *Hist Biol* **28**(5), 707-719.
37. **Han G, Liu Z, Liu X, Mao L, Jacques FMB., Wang X. 2016.** A whole plant herbaceous angiosperm from the Middle Jurassic of China. *Acta Geologica Sinica* **90**, 19-29.
38. **Liu Z-J, Wang X. 2016.** Yuhania: a unique angiosperm from the Middle Jurassic of Inner Mongolia, China. *Historical Biology*, 1-11.
39. **Dilcher DL., Sun G, Ji Q, Li HQ. 2007.** An early infructescence Hyrcantha decussata (comb. nov.) from the Yixian Formation in northeastern China. *Proceedings of the National Academy of Sciences of the United States of America* **104**(22), 9370-9374.
40. **Wang W, Dilcher DL, Sun G, Wang H-S, Chen Z-D. 2016.** Accelerated evolution of early angiosperms: Evidence from ranunculalean phylogeny by

integrating living and fossil data. *Journal of Systematics and Evolution* **54**(4), 336-341.
41. **Chang S-C., Zhang H, Hemming SR, Mesko GT, Fang Y. 2012.** Chronological evidence for extension of the Jehol Biota into Southern China. *Palaeogeography, Palaeoclimatology, Palaeoecology* **344–345**(0), 1-5.
42. **He HY, Wang XL, Zhou ZH, Wang F, Boven A, Shi G.H, Zhu RX. 2004.** Timing of the Jiufotang Formation (Jehol Group) in Liaoning, northeastern China, and its implications. *Geophysical Research Letters* **31**, 1-4.
43. **Jaramillo CA., Dilcher DL. 2001.** Middle Paleogene palynology of Central Colombia, South America: A study of pollen and spores from tropical latitudes. *Palaeontographica Abteilung B* **258**(4-6), 87-213.
44. **Germeraad JH., Hopping CA., Muller J. 1968.** Palynology of Tertiary sediments from tropical areas. *Review of Palaeobotany and Palynology* **6**, 189-348.
45. **Davis CC, Schaefer H, Xi Z, Baum D.A, Donoghue MJ, Harmon LJ. 2014.** Long-term morphological stasis maintained by a plant-pollinator mutualism. *Proceedings of the National Academy of Sciences of the United States of America* **111**(16), 5914-5919.
46. **Jaramillo CA, Rueda M, Torres V. 2010.** A palynological zonation for the Cenozoic of the Llanos and Llanos foothills of Colombia. *Palynology* **35**(1), 46-84.
47. **Boucher LD, Manchester SR, Judd WS. 2003.** An extinct genus of Salicaceae based on twigs with attached flowers, fruits, and foliage from the Eocene Green River Formation of Utah and Colorado, USA. *American Journal of Botany* **90**(9), 1389-1399.
48. **Manchester SR., Judd WS., Handley B. 2006.** Foliage and fruits of early poplars (Salicaceae: Populus) from the Eocene of Utah, Colorado, and Wyoming. *International Journal of Plant Sciences* **167**(4), 897-908.
49. **Smith ME, Carroll AR., Singer BS. 2008.** Synoptic revision of a major ancient lake system: Eocene Green River Formation, western United States. *GSA Bulletin* **120**, 54-84.
50. **Vandenberghe N, Hilgen FJ, Speijer RP. 2012.** The Paleogene Period. In *The geologic timescale 2012* (eds. Gradstein F.M., Ogg J.G., Schmitz M., Ogg G.), pp. 855-921. Ansterdam, Elsevier.
51. **Crepet WL, Nixon KC. 1998.** Fossil Clusiaceae from the Late Cretaceous (Turonian) of New Jersey and implications regarding the history of bee pollination. *American Journal of Botany* **85**(8), 1122-1133.
52. **Ruhfel BR, Stevens PF, Davis CC. 2013.** Combined Morphological and Molecular Phylogeny of the Clusioid Clade (Malpighiales) and the Placement of the Ancient Rosid Macrofossil Paleoclusia. *International Journal of Plant Sciences* **174**(6), 910-936.
53. **Massoni J, Doyle JA., Sauquet H. 2014.** Fossil calibration of Magnoliidae, an ancient lineage of angiosperms. *Palaeontologica Electronica*.
54. **Dettmann ME, Playford G. 1968.** Taxonomy of some Cretaceous spores and pollen grains from Eastern Australia. *Proceedings of the Royal Society of Victoria* **81**, 69-83.
55. **Sauquet H, Ho SY, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, Bayly MJ, Bromham L, Brown GK. Carpenter RJ, et al. 2012.** Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of Nothofagus (Fagales). *Syst Biol* **61**(2), 289-313.

56. **Partridge AD. 2006.** Late Cretaceous — Cenozoic palynology zonations Gippsland Basin. In *Australian Mesozoic and Cenozoic Palynology Zonations – updated to the 2004 Geologic Time Scale Geoscience Australia Record 2006/23* (ed. Monteil E), Geoscience Australia Record.
57. **Manchester SR, Dillhoff RM. 2004.** Fagus (Fagaceae) fruits, foliage, and pollen from the Middle Eocene of Pacific Northwestern North America. *Canadian Journal of Earth Sciences* **82**, 1509-1517.
58. **Renner SS., Grimm GW, Kapli P, Denk T. 2016.** Species relationships and divergence times in beeches: new insights from the inclusion of 53 young and old fossils in a birth-death clock model. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **371**(1699).
59. **Ewing TE. 1981.** Regional stratigraphy and structural setting of the Kamloops Group, South-central British Columbia. *Canadian Journal of Earth Sciences* **18**, 1464-1477.
60. **Moss PT, Greenwood DR, Archibald SB. 2005.** Regional and local vegetation community dynamics of the Eocene Okanagan Highlands (British Columbia – Washington State) from palynology. *Canadian Journal of Earth Sciences* **42**(2), 187-204.
61. **Pigg KB, DeVore Melanie L, Wojciechowski MF. 2008.** *Paleosecuridaca curtisii* gen. et sp. nov., Securidaca-like Samaras (Polygalaceae) from the Late Paleocene of North Dakota and their significance to the divergence of families within the Fabales. *International Journal of Plant Sciences* **169**(9), 1304-1313.
62. **Crane PR, Manchester SR, Dilcher DL 1990.** A preliminary survey of fossil leaves and well-preserved reproductive structures from the Sentinel Butte Formation (Paleocene) near Almont, North Dakota. *Fieldiana Geology* **20**(1418).
63. **Zetter R, Farabee MJ, Pigg KB, Manchester SR, DeVore ML., Nowak MD. 2011.** Palynoflora of the late Paleocene silicified shale at Almont, North Dakota, USA. *Palynology* **35**(2), 179-211.
64. **McClain AM, Manchester SR. 2001.** Dipteronia (Sapindaceae) from the Tertiary of North America and Implications for the Phytogeographic History of the Aceroideae. *American Journal of Botany* **88**(7), 1316-1325.
65. **Qiu YL, Li LB, Wang B, Chen ZD, Dombrovska O, Lee J, Kent L, Li RQ, Jobson RW, Hendry TA, et al. 2007.** A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *International Journal of Plant Sciences* **168**(5), 691-708.
66. **Manchester SR, Chen Z-D, Lu A-M, Uemura K. 2009.** Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern Hemisphere. *Journal of Systematics and Evolution* **47**(1), 1-42.
67. **Nichols DJ, Ott HL. 1978.** Biostratigraphy and evolution of the *Momipites-Caryapollenites* lineage in the early Tertiary in the Wind River Basin, Wyoming. *Palynology* **2**, 93-112.
68. **Peppe DJ. 2010.** Megafloral change in the early and middle Paleocene in the Williston Basin, North Dakota, USA. *Palaeogeography, Palaeoclimatology, Palaeoecology* **298**, 224-234.
69. **Corbett SL, Manchester SR. 2004.** Phytogeography and Fossil History of Ailanthus (Simaroubaceae). *International Journal of Plant Sciences* **165**(4), 671-690.

70. **Takahashi M, Crane PR, Ando H. 1999.** *Esgueiria futabensis* sp. nov., a new angiosperm flower from the Upper Cretaceous (lower Coniacian) of northeastern Honshu, Japan. *Paleontological Research* **3**, 81-87.
71. **Friis EM., Pedersen KR., Crane PR. 2006.** Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. *Palaeogeography, Palaeoclimatology, Palaeoecology* **232**(2-4), 251-293.
72. **Yazykova E. 2002.** Ammonite and inoceramid radiations after the Santonian–Campanian bioevent in Sakhalin, Far East Russia. *Lethaia* **35**, 51-60.
73. **Manchester SR, Kapgate DK, Wen J. 2013.** Oldest fruits of the grape family (Vitaceae) from the Late Cretaceous Deccan Cherts of India. *American Journal of Botany* **100**(9), 1849-1859.
74. **Schoene B, Samperton KM, Eddy MP., Keller G, Adatte T, Bowring SA, Khadri SFR., Gertsch B. 2015**. U-Pb geochronology of the Deccan Traps and relation to the end-Cretaceous mass extinction. *Science* **347**, 182-184.
75. **Hermsen EJ, Gandolfo MA, Nixon K.C, Crepet WL. 2003** *Divisestylus* gen. nov. (aff. Iteaceae), a fossil saxifrage from the Late Cretaceous of New Jersey, USA. *American Journal of Botany* **90**, 1371-1388.
76. **Magallón S, Herendeen PS., Crane PR. 2001**. Androdecidua endressii gen. et sp. nov., from the Late Cretaceous of Georgia (United States): Further Floral Diversity in Hamamelidoideae (Hamamelidaceae). *International Journal of Plant Sciences* **162**(4), 963-983.
77. **Sims HJ, Herendeen PS., Crane PR. 1998.** New Genus of Fossil Fagaceae from the Santonian (Late Cretaceous) of Central Georgia, U. S. A. *International Journal of Plant Sciences* **159**(2), 391-404.
78. **Manchester SR, Donoghue MJ. 1995.** Winged fruits of Linnaeeae (Caprifoliaceae) in the Tertiary of western North America: *Diplodipelta* gen. nov. *International Journal of Plant Sciences* **156**(5), 709-722.
79. **Bell CD, Donoghue MJ. 2005.** Dating the dipsacales: Comparing models, genes, and evolutionary implications. *American Journal of Botany* **92**(2), 284-296.
80. **Evanoff E, McIntosh WC, Murphey PC. 2001.** Stratigraphic summary and 40Ar/39Ar geochronology of the Florissant Formation, Colorado. In *Fossil flora and stratigraphy of the florissant Formation, Colorado: Proceedings of the Denver Museum of Nature & Science, scr 4, no I, p 1-16* (eds. Evanoff E, Gregory-Wodzicki KM. Johnson KR), pp. 1-16. Denver, Denver Museum of Nature & Science.
81. **Prothero DR. 2008.** Magnetic stratigraphy of the Eocene-Oligocene floral transition in western North America. In Meyer, ll. W., and Smith, D. M., eds., Paleomology of the Upper Eocene Florissant Formation, Colorado. *Geological Society of America Special Paper 435*, 71-87.
82. **Martínez-Millán M. 2010.** Fossil record and age of the Asteridae. *Botanical Review* **76**, 83-135.
83. **Taylor D.W. 1989.** Select palynomorphs from the Middle Eocene Claiborne Formation, Tenn., (U.S.A.). *Review of Palaeobotany and Palynology* **58**(2-4), 111-128.
84. **Elsik WC. 1974.** Characteristic Eocene palynomorphs in the Gulf Coast, U.S.A. *Palaeontographica Abteilung B* **149**, 90-111.
85. **Zavada M, de Villiers S. 2000.** Pollen of the Asteraceae from the Paleocene-Eocene of South Africa. *Grana* **39**(1), 39-45.
86. **Wade BS., Pearson PN, Berggren WA, Pälike H. 2011.** Review and revision of Cenozoic tropical planktonic foraminiferal biostratigraphy and calibration to

the geomagnetic polarity and astronomical time scale. *Earth-Science Reviews* **104**(1–3), 111-142.
87. **Mai DH. 1970.** Neue Arten aus tertiären Lorbeerwäldern in Mitteleuropa (Vorläufige Mitteilung). *Feddes Repertorium* **98**, 347-370.
88. **Mai DH. 1987.** Neue Früchte und Samen aus paläozänen Ablagerungen Mitteleuropas. *Feddes Repertorium* **98**, 197-229.
89. **Knobloch EN, Mai DH. 1986.** *Monographie der Früchte und Samen in der Kreide von Mitteleuropa*. Prague, Ústřední ústav geologický v Academii, nakl. Československé akademie věd.
90. **Knobloch E, Kvaček Z, Bůžek Č, Mai DH, Batten DJ. 1993.** Evolutionary significance of floristic changes in the Northern Hemisphere during the Late Cretaceous and Palaeogene, with particular reference to Central Europe. *Review of Palaeobotany and Palynology* **78**(1–2), 41-54.
91. **Berry EW. 1930.** Revision of the Lower Eocene Wilcox Flora of the southeastern states with descriptions of new species, chiefly from Tennessee and Kentucky. *United States Geological Survey Professional Paper* **156**, 1-196.
92. **Millan M, Crepet WL. 2014**. The fossil record of the Solanaceae revisited and revised—the fossil record of Rhamnaceae enhanced. *Botanical Review* **80**, 73-106.
93. **Crepet WL., Nixon KC, Daghlian CP. 2012.** Fossil Ericales from the Upper Cretceous of New Jersey. *International Journal of Plant Sciences* **174**, 572-584.
94. **Gandolfo MA., Nixon KC., Crepet WL. 1988.** *Tylerianthus crossmanensis* gen. et sp. nov. (Rosales) from the Upper Cretaceous of New Jersey. *American Journal of Botany* **85**, 376-386.
95. **Gandolfo MA, Nixon KC., Crepet WL 1998.** A new fossil flower from the Turonian of New Jersey: Dressiantha bicarpellata gen. et sp. nov. (Capparales). *American Journal of Botany* **85**(7), 964-974.
96. **Crepet WL., Nixon KC., Gandolfo MA. 2004.** Fossil evidence and phylogeny: The age of major angiosperm clades based on mesofossil and macrofossil evidence from Cretaceous deposits. *American Journal of Botany* **91**(10), 1666-1682.
97. **Drinnan AN, Crane PR, Friis EM, Pedersen KR. 1991.** Angiosperm flowers and tricolpate pollen of buxaceous affinity from the Potomac Group (mid-Cretaceous) of eastern North America. *American Journal of Botany* **78**, 153-176.
98. **Doyle JA. 2015.** Recognising angiosperm clades in the Early Cretaceous fossil record. *Historical Biology* **27**(3-4), 414-429.
99. **Hochuli PA., Heimhofer U, Weissert H. 2006.** Timing of early angiosperm radiation: recalibrating the classical succession. *Journal of the Geological Society* **163**, 587-594.
100. **Crane PR, Pedersen KR, Friis EM, Drinnan AN. 1993**. Early Cretaceous (early to middleAlbian) platanoid inflorescences associated with Sapindopsis leaves from the Potomac Group of eastern North America. *Systematic Botany* **18**, 328-344.
101. **Jacques FMB., Wang W, CORD, Li H-L., Zhou Z-K., Chen Z-D. 2010.** Integrating fossils in a molecular-based phylogeny and testing them as calibration points for divergence time estimates in Menispermaceae. *Journal Systematics and Evolution* **49**, 25-49.

102. **Pacltová B. 1981**. The evolution and distribution of *Normapolles* pollen during the Cenophytic. *Review of Palaeobotany & Palynology* **35**, 175-208.
103. **Knobloch E. 1985.** Paläobotanisch-biostratigraphische Charakteristik der Klikov-Schichtenfolge (Oberturon–Santon) in Südböhmen. *Sborník geologických věd, Geologie* **40**, 101–145.
104. **Doyle JA, Endress PK, Upchurch GR. Jr. 2008.** Early Cretaceous monocots: a phylogenetic evaluation. *Sbornik Narodniho Muzea v Praze Rada B Prirodni Vedy* **64**(2-4), 59-87.
105. **Hughes NF. 1994.** *The enigma of angiosperm origins*. Cambridge, Cambridge University Press.
106. **Hertweck KL, Kinney MS, Stuart SA, Maurin O, Mathews S, Chase MW, Gandolfo MA, Pires JC. 2015.** Phylogenetics, divergence times and diversification from three genomic partitions in monocots. *Botanical Journal of the Linnean Society* **178**(3), 375-393.
107. **Friis EM. 1988.** *Spirematospermum chandlerae* sp. nov., an extinct species of Zingiberacea from the North American Cretaceous. *Tertiary Research* **9**, 7-12.
108. **Christopher RA, Prowell DC 2010.** A palynological biozonation for the uppermost Santonian and Campanian Stages (Upper Cretaceous) of South Carolina, USA. *Cretaceous Research* **31**(2), 101-129.
109. **Berry EW. 1914.** The Upper Cretaceous and Eocene floras of South Carolina and Georgia. *United States Geological Survey Professional Paper* **84**, 1-200.
110. **Iles WJD, Smith SY, Gandolfo MA, Graham SW. 2015.** Monocot fossils suitable for molecular dating analyses. *Botanical Journal of the Linnean Society* **178**(3), 346-374.
111. **Sohl NF, Owens JP. 1991.** Cretaceous stratigraphy of the Carolina Coastal Plain. In *The Geology of the Carolinas: Knoxville* (eds. Horton J.W., Jr., Zullo V.A.), pp. 191-220, University of Tennessee Press.
112. **Campbell BG, Gohn GS. 1994.** Stratigraphic framework for geologic and geohydrologic studies of the subsurface Cretaceous section near Charleston, South Carolina. *United States Geological Survey Map MF-2273*, 1-11.
113. **Habib D, Miller JA. 1989.** Dinoflagellate species and organic facies evidence of marine transgression and regression in the atlantic coastal plain. *Palaeogeography, Palaeoclimatology, Palaeoecology* **74**(1–2), 23-47.
114. **Ramirez SR, Gravendeel B, Singer RB., Marshall CR., Pierce NE. 2007**. Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature* **448**(7157), 1042-1045.
115. **Iturralde-Vinent MA., MacPhee RDE. 1996.** Age and paleogeographical origin of Dominican Amber. *Science* **273**, 1850-1852.
116. **Conran JG., Bannister JM, Mildenhall DC, Lee DE, Chacon J, Renner SS. 2014.** Leaf fossils of Luzuriaga and a monocot flower with in situ pollen of Liliacidites contortus Mildenh. & Bannister sp. nov. (Alstroemeriaceae) from the Early Miocene. *Am J Bot* **101**(1), 141-155
117. **Lindqvist JK, Lee DE. 2009.** High-frequency paleoclimate signals from Foulden Maar, Waipiata Volcanic Field, southern New Zealand: an early Miocene varved lacustrine diatomite deposit. *Sedimentary Geology* **222**, 98-110.
118. **Mildenhall DC., Kennedy EM, Lee DE, Kaulfuss U, Bannister JM, Fox B, Conran JG. 2014.** Palynology of the early Miocene Foulden Maar, Otago,

New Zealand: Diversity following destruction. *Review of Palaeobotany and Palynology* **204**(0), 27-42.
119. **Cooper RA. 2004.** The New Zealand geological timescale. *Institute of Geological and Nuclear Sciences Monograph* **22**, 1-284.
120. **Hilgen FJ., Lourens LJ, Van Dam JA. 2012.** The Neogene Period. In *The Geologic Times Scale 2012* (eds. Gradstein F.M., Ogg J.G., Schmitz M., Ogg G.), pp. 923-978, Elsevier.
121. **Gandolfo MA., Nixon K.C, Crepet WL. 2002.** Triuridaceae fossil flowers from the Upper Cretaceous of New Jersey. *American Journal of Botany* **89**(12), 1940-1957.
122. **Friis EM, Pedersen KR, Crane PR. 2004.** Araceae from the Early Cretaceous of Portugal: Evidence on the emergence of monocotyledons. *Proceedings of the National Academy of Sciences of the United States of America* **101**(47), 16565-16570.
123. **Friis EM, Pedersen KR, Crane PR. 2010.** Diversity in obscurity: fossil flowers and the early history of angiosperms. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 369-382.
124. **Dinis JL, Rey J, Cunha PP, Callapez P, dos Reis RP. 2008.** Stratigraphy and allogenic controls of the western Portugal Cretaceous: an updated synthesis. *Cretaceous Research* **29**(5-6), 772-780.
125. **Berthou PY. 1973.** Le Cénomanien de l'Estremadure portugaise. *Memória, Serviços Geológicos de Portugal, Lisboa, NS* **23**, 169.
126. **Berthou PY. 1984.** Résumé synthetique de la stratigraphie et de la paléogéographie du Cretacé moyen et supérieur du bassin occidental portugais. *Geonovas* **7**, 99-120.
127. **Riley MG., Stockey RA. 2004**. *Cardstonia tolmanii* gen. et sp. nov. (Limnocharitaceae) from the Upper Cretaceous of Alberta, Canada. *International Journal of Plant Sciences* **165**(5), 897-916.
128. **Eberth DA., Braman DR. 2012.** A revised stratigraphy and depositional history for the Horseshoe Canyon Formation (Upper Cretaceous), southern Alberta plains. *Canadian Journal of Earth Sciences* **49**, 1053-1086.
129. **Bogner J, Johnson KR, Kvacek Z, Upchurch Jr. GR. 2007.** New fossil leaves of Araceae from the Late Cretaceous and Paleogene of western North America. *Zitteliana* **A47**, 133-147.
130. **Kvaček J, Herman AB. 2004.** Monocotyledons from the Early Campanian (Cretaceous) of Grünbach, Lower Austria. *Review of Palaeobotany and Palynology* **128**(3–4), 323-353.
131. **Nauheimer L, Metzler D, Renner SS. 2012.** Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. *New Phytologist* **195**(4), 938-950.
132. **Herman AB, Kvaček J. 2010**. *Late Cretaceous Grünbach Flora of Austria*. Vienna, Naturhistorisches Museum Wien.
133. **Friis EM, Eklund H, Pedersen KR., Crane PR. 1994.** *Virginianthus calycanthoides* gen. et sp. nov.—A calycanthaceous flower from the Potomac Group (Early Cretaceous) of eastern North America. *International Journal of Plant Sciences* **155**, 772-785.
134. **Doyle JA., Endress PK. 2010.** Integrating Early Cretaceous fossils into the phylogeny of living angiosperms: Magnoliidae and eudicots. *Journal of Systematics and Evolution* **48**(1), 1-35.

135. **von Balthazar M, Crane PR, Pedersen KR, Friis EM. 2011.** New flowers of Laurales from the Early Cretaceous (Early to Middle Albian) of eastern North America. In *Flowers on the tree of life* (eds. Wanntorp L., Ronse De Craene L.P.), pp. 49–87. Cambridge Cambridge University Press.
136. **Doyle JA, Robbins EI. 1977.** Angiosperm pollen zonation of the continental Cretaceous of the Atlantic coastal plain and its application to deep wells in the Salisbury Embayment. *Palynology* (1), 43-78.
137. Mohr B.A.R., Bernardes-de-Oliveira M.E.C. 2004 Endressinia brasiliana, a magnolialean angiosperm from the lower Cretaceous Crato Formation (Brazil). *International Journal of Plant Sciences* **165**(6), 1121-1133.
138. **Mohr BAR, Coiffard C, Bernardes-de-Oliveira MEC. 2013.** Schenkeriphyllum glanduliferum, a new magnolialean angiosperm from the Early Cretaceous of Northern Gondwana and its relationships to fossil and modern Magnoliales. *Review of Palaeobotany and Palynology* **189**, 57-72.
139. **Sauquet H, Doyle JA, Scharaschkin T, Borsch T, Hilu K.W, Chatrou L.W, Le Thomas A. 2003.** Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society* **142**, 125-186.
140. **Batten DJ. 2007.** Spores and pollen from the Crato Formation: biostratigraphic and palaeoenvironmental implications. In *The Crato fossil beds of Brazil - window into an ancient world* (eds. Martill D.M., Bechly G., Loveridge R.F.), pp. 566-573. Cambridge, Cambridge University Press.
141. **Martill DM. 2007.** The age of the Cretaceous Santana Formation fossil Konservat Lagerstatten of north-east Brazil: a historical review and an appraisal of the biochronostratigraphic utility of its palaeobiota. *Cretaceous Research* **28**(6), 895-920.
142. **Heimhofer U, Hochuli P-A. 2010.** Early Cretaceous angiosperm pollen from a low-latitude succession (Araripe Basin, NE Brazil). *Review of Palaeobotany & Palynology* **161**(3-4), 105-126.
143. **Friis EM., Pedersen KR., Crane PR. 1995.** *Appomattoxia ancistrophora* gen et sp nov, a new Cretaceous plant with similarities to Circaeaster and extant Magnoliidae. *American Journal of Botany* **82**(7), 933-943.
144. **Friis E.M, Pedersen KR., Crane PR. 2010.** Cretaceous diversification of angiosperms in the western part of the Iberian Peninsula. *Review of Palaeobotany and Palynology* **162**(3), 341-361.
145. **Smith SY, Stockey A. 2007.** Establishing a fossil record for the perianthless Piperales: *Saururus tuckerae* sp. nov. (Saururaaceae) from the Middle Eocene Princeton Chert. *American Journal of Botany* **94**, 1642–1657.
146. **Rouse GE, Mathews W. 1961,** Radioactive dating of Tertiary plant-bearing deposits. *Science* **133**, 1079-1080.
147. **Mathews WH. 1964.** Potassium-argon age determinations of Cenozoic volcanic rocks from British Columbia. *Geological Society of America Bulletin* **75**, 465-468.
148. **Hills LV, Baadsgaard H. 1967.** Potassium-argon dating of some Lower Tertiary strata in British Columbia. *Bulletin of Canadian Petroleum Geology* **15**, 138-149.
149. **Read PB. 2000.** Geology and industrial minerals of the Tertiary basins, south-central British Columbia. *British Columbia Geological Survey Geo-File* **2000**(3).
150. **Moss P.T, Greenwood DR., Archibald SB. 2005.** Regional and local vegetation community dynamics of the Eocene Okanagan Highlands (British

Columbia – Washington State) from palynology. *Canadian Journal of Earth Sciences* **42**, 187-204.
151. **Hughes NF, Drewry G, Laing J.F. 1979.** Barremian earliest angiosperm pollen. *Palaeontology* **22**, 513-535.
152. **Friis EM., Pedersen KR., Crane PR. 2000.** Fossil floral structures of a basal angiosperm with monocolpate, reticulate-acolumellate pollen from the Early Cretaceous of Portugal. *Grana* **39**, 226–239.
153. **Doyle JA., Endress PK. 2014.** Integrating Early Cretaceous Fossils into the Phylogeny of Living Angiosperms: ANITA Lines and Relatives of Chloranthaceae. *International Journal of Plant Sciences* **175**(5), 555-600.
154. **Rocha RB., Manuppella G, Mouterde R, Ruget C, Zbyszewski G. 1981.** *Carta geológica de Portugal na escala de 1/50 00. Notícia explicativa da folha 19-C Figueria da Foz*. Lisboa, Serviços Geológicos de Portugal.
155. **Manuppella G, Telles Antunes M, Costa Almedia CA, Azeredo AC, Barbosa B, Cardosa JL., Crispim JA., Duarte LV., Henriques MH., Martins LT., et al. 2000.** *Carta geoógica de Portugal na escala de 1:50 000. Notícia explicativa da folha 27-A (Vila Nova de Ourém)*. Lisboa, Instituto Geológico e Mineiro; 156 p.
156. **Dinis JL. 2001.** Definicão da Formacão da Figueira da Foz-Aptiano a Cenomaniano do sector central da margem oesteibé rica. *Comunicações Instituto Geológico e Mineiro* **88**, 127– 160.
157. **Callapez P. 2003.** The Cenomanian-Turonian transition in West Central Portugal: ammonites and biostratigraphy. *Ciências da Terra* **15**, 53-70.
158. **Frumin S, Friis EM. 1999.** Magnoliid reproductive organs from the Cenomanian-Turonian of north-western Kazakhstan: Magnoliaceae and Illiciaceae. *Plant Systematics and Evolution* **216**(3-4), 265-288.
159. **Levina AP, Zhelezko VI, Leiptsig AB, Papulov GN, Ponomarenko ZK, Paskar ZS. 1990.** The Sokolov and Sarbay ironmine quarries. In *Upper Cretaceous deposits of the southern Transural (The region of the upper current of the Tobol River)* (eds. Papulov G.N., Zhelezko V.I., Levina A.R.), pp. 46-58. Sverdlovsk, Akademia Nauk SSSR.
160. **Korchagin OA, Bragina LG., Bragin NY. 2012.** Planktonic foraminifers and radiolarians from the Coniacian-santonian deposits of the Mt. Ak-Kaya, Crimean Mountains, Ukraine. *Stratigraphy and Geological Correlation* **20**(1), 73-96.
161. **Mohr BAR., Bernardes-De-Oliveira MEC, Taylor DW. 2008.** Pluricarpellatia, a nymphaealean angiosperm from the Lower Cretaceous of northern Gondwana (Crato Formation, Brazil). *Taxon* **57**(4), 1147-1158.
162. **Taylor DW., Brenner GJ., Basha SH. 2008.** Scutifolium jordanicum gen. et sp nov (Cabombaceae), an aquatic fossil plant from the Lower Cretaceous of Jordan, and the relationships of related leaf fossils to living genera. *American Journal of Botany* **95**(3), 340-352.
163. **Friis E.M, Pedersen KR., Crane PR. 2001.** Fossil evidence of water lilies (Nymphaeales) in the Early Cretaceous. *Nature* **410**(6826), 357-360.
164. **Friis EM, Pedersen KR, Von Balthazar M, Grimm G.W, Crane PR. 2009**. Monetianthus mirus gen. et sp. nov., a nymphaealean flowers from the Early Cretaceous of Portugal. *International Journal of Plant Science* **170**, 1086-1101.
165. **Smith SA., Beaulieu JM., Donoghue MJ. 2010.** An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences* **107**(13), 5897-5902.

166. **Zanne AE., Tank DC., Cornwell WK, Eastman JM., Smith SA., FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT., Reich PB., et al. 2014**. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**(7486), 89-92.
167. **Amireh BS., Jarrar G, Henjes-Kunst F, Schneider W. 1998**. K-Ar dating, X-ray diffractometry, optical and scanning electron microscopy of glauconites from the Early Cretaceous Kurnub. *Geological Journal* **33**, 49-65.
168. **Wieland GW. 1935.** The Cerro Cuadrado petrified forest. *Carnegie Institution of Washington Publication* **449**, 1-183.
169. **Calder MG. 1953.** A coniferous petrified forest in Patagonia. *Bulletin of the British Museum (Natural History): Geology* **2**, 99-138.
170. **Stockey RA. 1975.** Seeds and embryos of *Araucaria mirabilis*. *American Journal of Botany* **62**(8), 856-868.
171. **Stockey RA. 1978.** Reproductive biology of Cerro Cuadrado fossil conifers: Ontogeny and reproductive strategies in *Araucaria mirabilis* (Spegazzini) Windhausen. *Palaeontographica Abteilung B* **166**, 1-15.
172. **Wilde MH., Eames AJ. 1948.** The ovule and seed of *Araucaria badwillii* with discussion of the taxonomy of the genus. 1. Morphology. *Annals of botany* **12**(47), 311-&.
173. **Spalleti L, Iñiguez Rodríguez AM, Masón M. 1982.** Edades radimétricas de piroclastitas y volvanitas del Grupo Bahía Laura, Gran Bajo de San Julián, Santa Cruz. *Revista de la Asociación Geológica Argentina* **37**, 483-485.
174. **Florin R. 1951.** Evolution in cordaites and conifers. *Acta Horti Bergiani* **15**, 285-388.
175. Yao X.L., Taylor T.N., Taylor E.L. 1997 A taxodiaceous seed cone from the Triassic of Antarctica. *American Journal of Botany* **84**(3), 343-354.
176. **Axsmith BJ., Taylor TN, Taylor EL. 1998.** Anatomically preserved leaves of the conifer Notophytum krauselii (Podocarpaceae) from the Triassic of Antarctica. *American Journal of Botany* **85**(5), 704-713.
177. **Krassilov VA. 1986.** New floral structure from the Lower Cretaceous of Lake Baikal Area. *Review of Palaeobotany and Palynology* **47**, 9-16.
178. **Rydin C, Pedersen KR, Friis EM. 2004.** On the evolutionary history of *Ephedra*: Cretaceous fossils and extant molecules. *Proceedings of the National Academy of Sciences, USA* **101**, 16571-16576.
179. **Friis EM., Pedersen KR., Crane PR. 2009.** Early Cretaceous mesofossils from Portugal and eastern North America related to the Bennettitales-Erdtmanithecales-Gnetales group. *American Journal of Botany* **96**, 252-283.
180. **Godefroit P. 2012**. *Bernissart dinosaurs and Early Cretaceous terrestrial ecosystems*. Bloomington, IN, Indiana University Press; 648 p.
181. **Vakhrameev V, Kotova I. 1977.** Ancient angiosperms and accompanying plants from the Lower Cretaceous of Transbaikalia. *Paleontological Journal* **4**, 487-495.
182. **Vakhrameev V. 1991**. *Jurassic and Cretaceous floras and climates of the Earth*. Cambridge, Cambridge University Press; 318 p.
183. **Chen P, Wang Q, Zhang H, Cao M, Li W, Wu S, Shen Y. 2005.** Jianshangou Bed of the Yixian Formation in West Liaoning, China. *Science in China Series D: Earth Sciences* **48**(3), 298-312.
184. **Dettmann ME., Thomson MRA. 1987.** Cretaceous palynomorphs from the James-Ross Island area, Antarctica - a pilot-study. *British Antarctic Survey Bulletin* **77**, 13-59.

185. **Gao Z, Thomas BA. 1989.** A review of fossil cycad megasporophylls, with new evidence of Crossozamia pomel and its associated leaves from the lower Permian of Taiyuan, China. *Review of Palaeobotany and Palynology* **60**, 205-223.
186. **Nagalingum NS, Marshall CR, Quental TB, Rai HS, Little DP., Mathews S. 2011.** Recent synchronous radiation of a living fossil. *Science* **334**(6057), 796-799.
187. **Hermsen EJ., Taylor TN., Taylor EL, Stevenson DW. 2006.** Cataphylls of the Middle Triassic cycad Antarcticycas schopfii and new insights into cycad evolution. *American Journal of Botany* **93**(5), 724-738.
188. **Wang J. 2010.** Late Paleozoic macrofloral assemblages from Weibei Coalfield, with reference to vegetational change through the Late Paleozoic Ice-age in the North China Block. *International Journal of Coal Geology* **83**(2-3), 292-317.
189. **Henderson CM, Gradstein FM., Hammer O. 2012.** The Permian Period. In *The geologic timescale 2012* (eds. Gradstein F.M., Ogg J.G., Schmitz M., Ogg G.), pp. 653-679, Elsevier.
190. **Wellman CH, Gensel PG, Taylor WA. 2009.** Spore wall ultrastructure in the early lycopsid *Leclercqia* (Protolepidodenrales) from the Lower Devonian of North America: evidence for a fundamental division in the lycopsids. *American Journal of Botany* **96**(10), 1849-1860.
191. **Meyer-Berthaud B, Fairon-Demaret M, Steemans P, Talent J, Gerrienne P. 2003**. The plant Leclercqia (Lycopsida) in Gondwana: implications for reconstructing Middle Devonian palaeogeography. *Geological Magazine* **140**(2), 119-130.

B. Supplementary figures and tables for chapter three

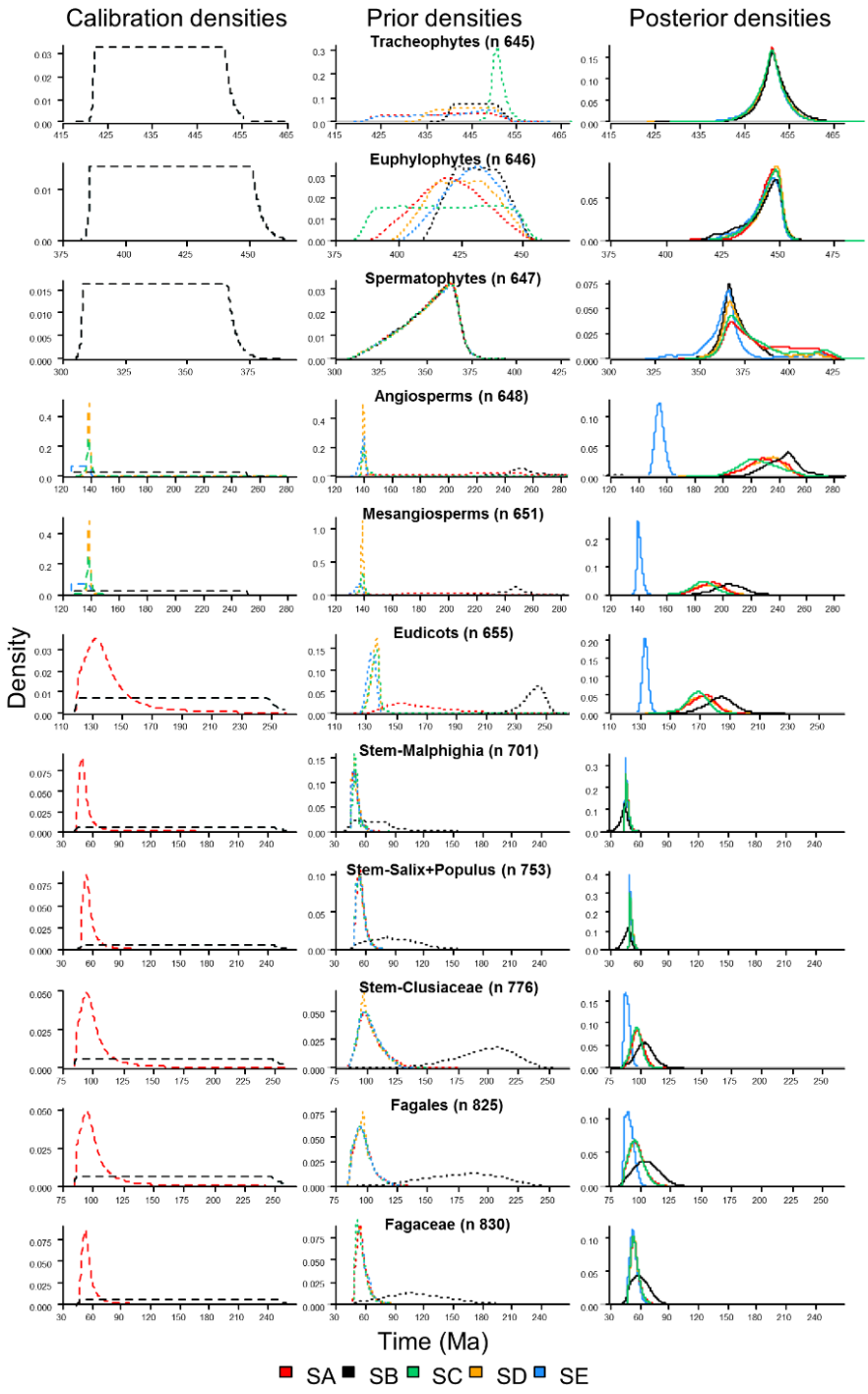


Figure B.1: Calibration, prior and posterior densities for 52 calibrated nodes in the tree and for the 5 calibration strategies. Fossil calibrations are shown as dashed lines, priors as doted lines and posterior as solid lines. Nodes are numbered as in Figure B.2. Estimates were obtained using the HKY85+Γ5 substitution model, independent rates model, with the 83 genes subdivided into three partitions: (1) 1st and 2nd codon positions for plastid genes; (2) 1st and 2nd codon positions for mitochondrial genes; and (3) nuclear RNA genes. Colouring relates to the calibration strategy (SA, SB, SC, SD, SE) as in Figure 3.3.

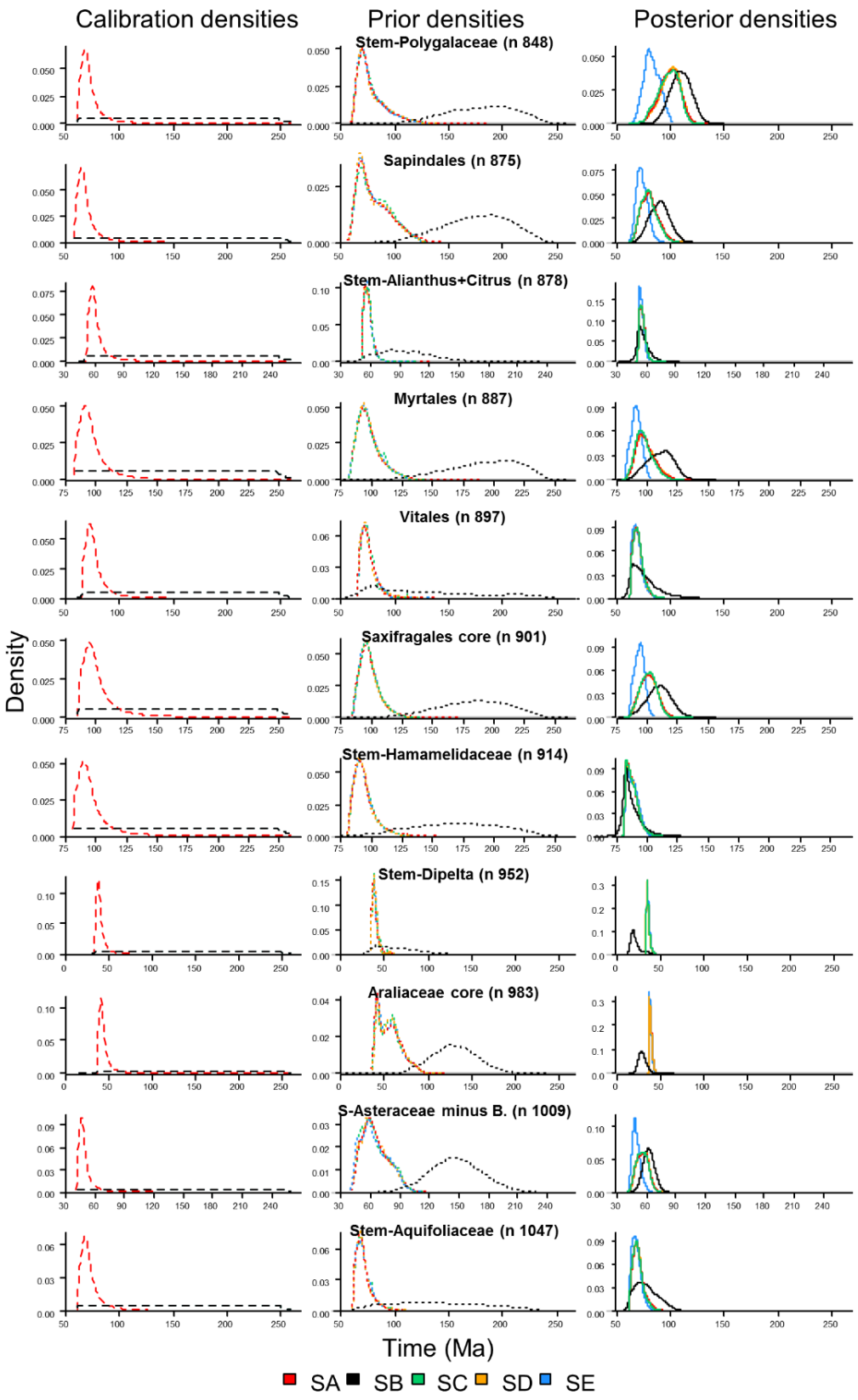


Figure B.1.1: Calibration, prior and posterior densities for 52 calibrated nodes in the tree and for the 5 calibration strategies. Continued.

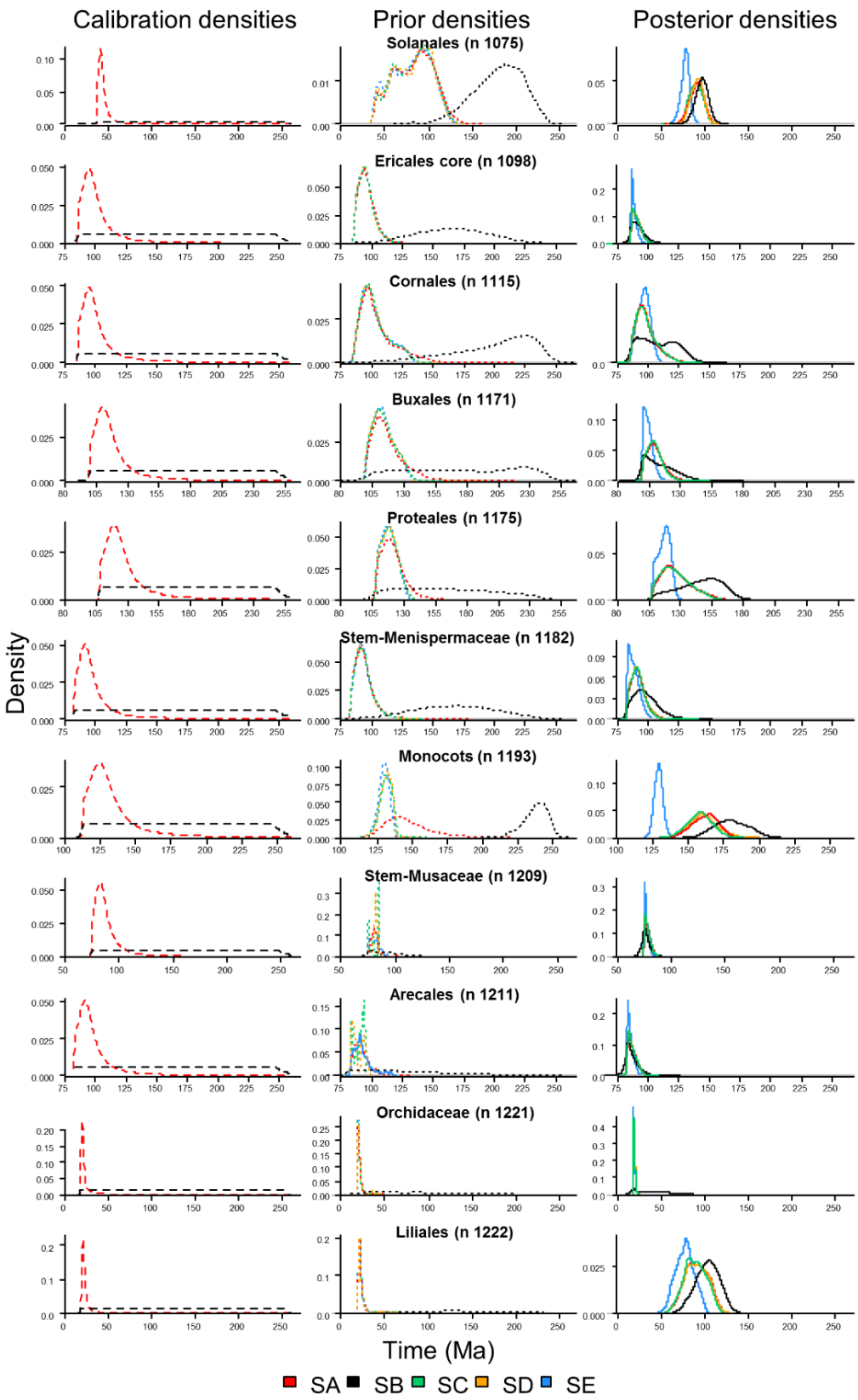


Figure B.1.2: Calibration, prior and posterior densities for 52 calibrated nodes in the tree and for the 5 calibration strategies. Continued.

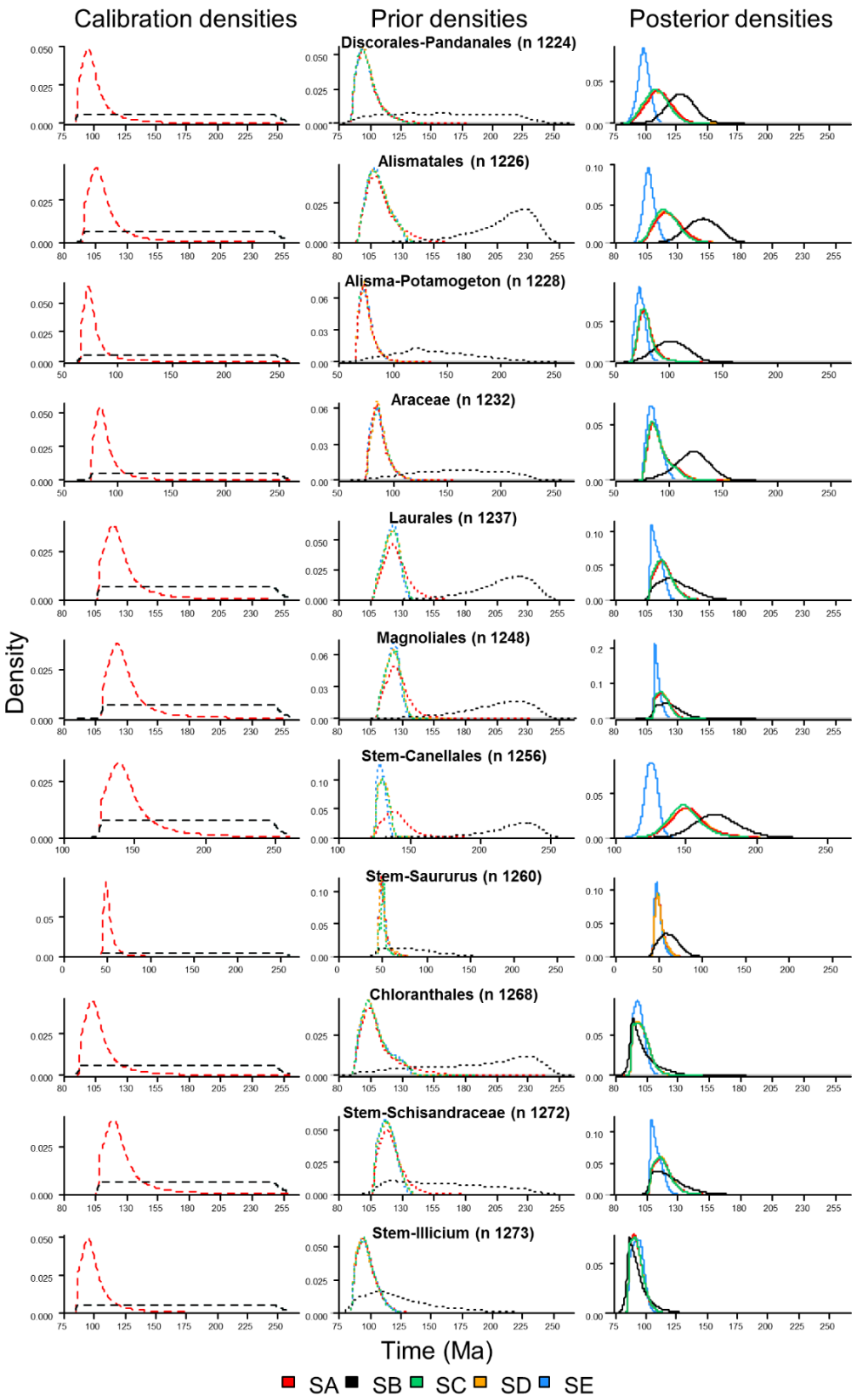


Figure B.1.3: Calibration, prior and posterior densities for 52 calibrated nodes in the tree and for the 5 calibration strategies. Continued.

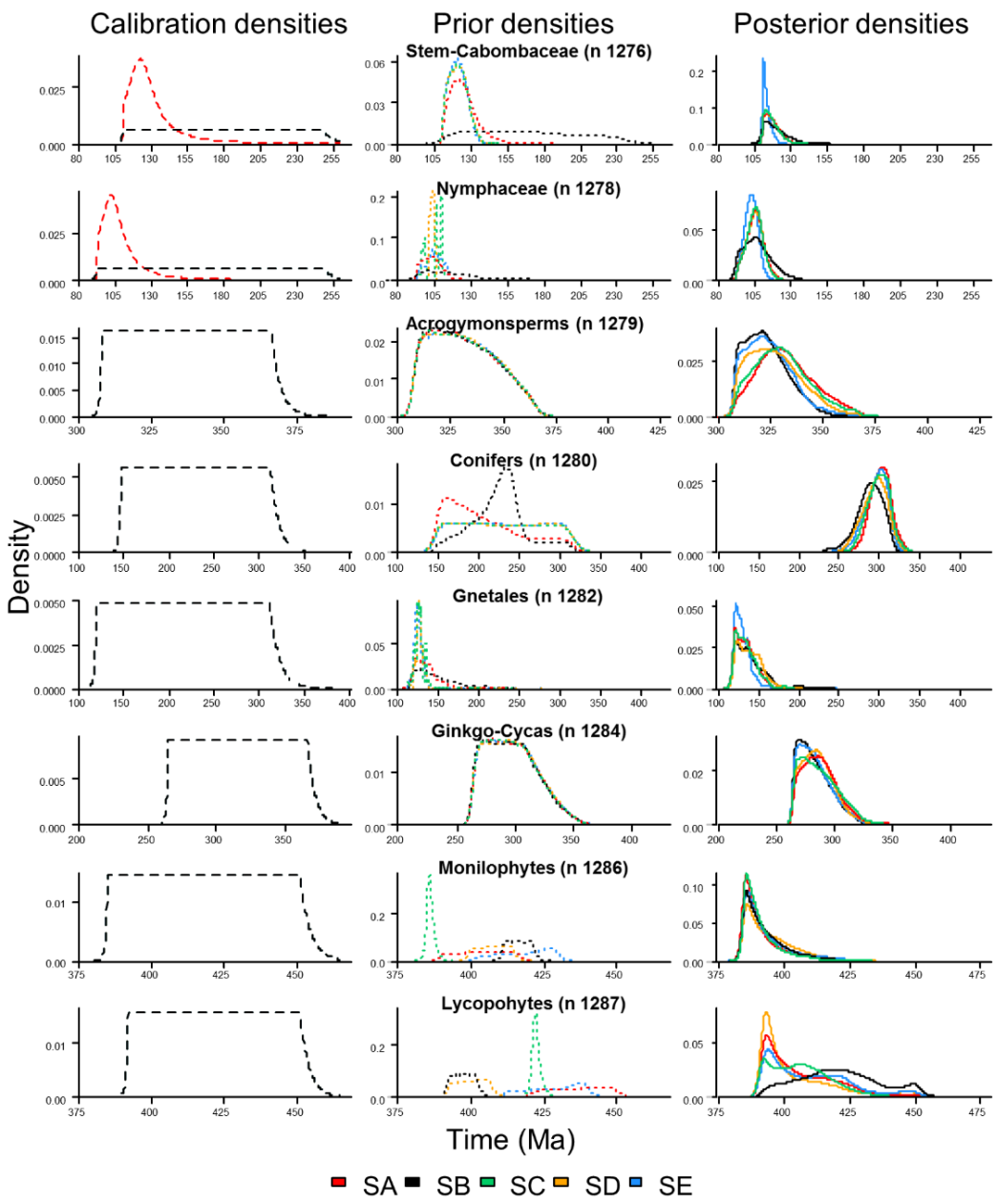


Figure B.1.4: Calibration, prior and posterior densities for 52 calibrated nodes in the tree and for the 5 calibration strategies. Continued.

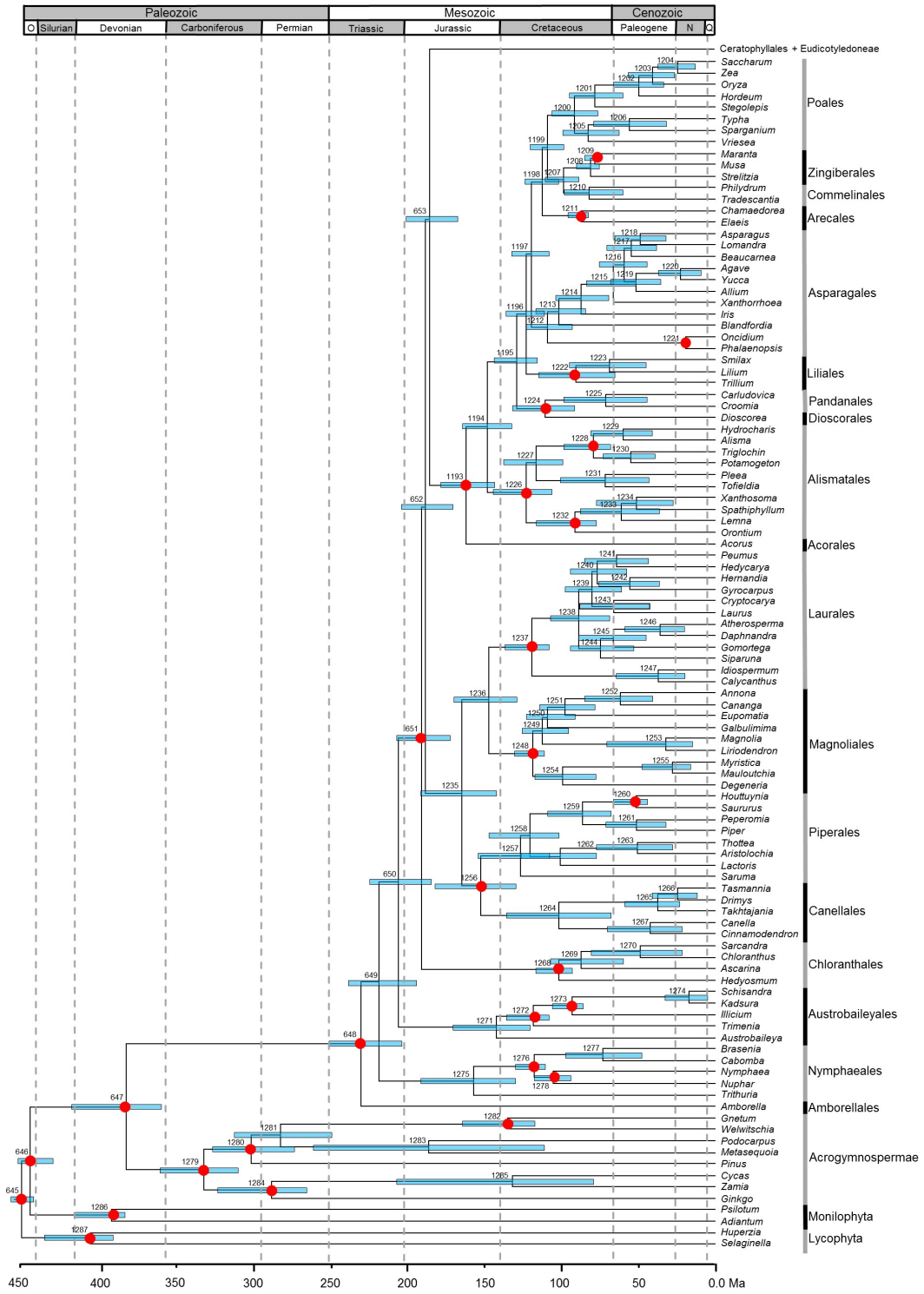


Figure B.2: Chronogram of 644 taxa of tracheophytes (from SA-IR-3P). Blue bars represent the 95% HPD credibility intervals for the node ages. Red dots represent calibrated nodes. Divergence times were estimated using calibration strategy A, HKY85+Γ5 substitution model, independent rates model, with the 83 genes subdivided into three partitions: (1) 1st and 2nd codon positions for plastid genes; (2) 1st and 2nd codon positions for mitochondrial genes; and (3) nuclear RNA genes.

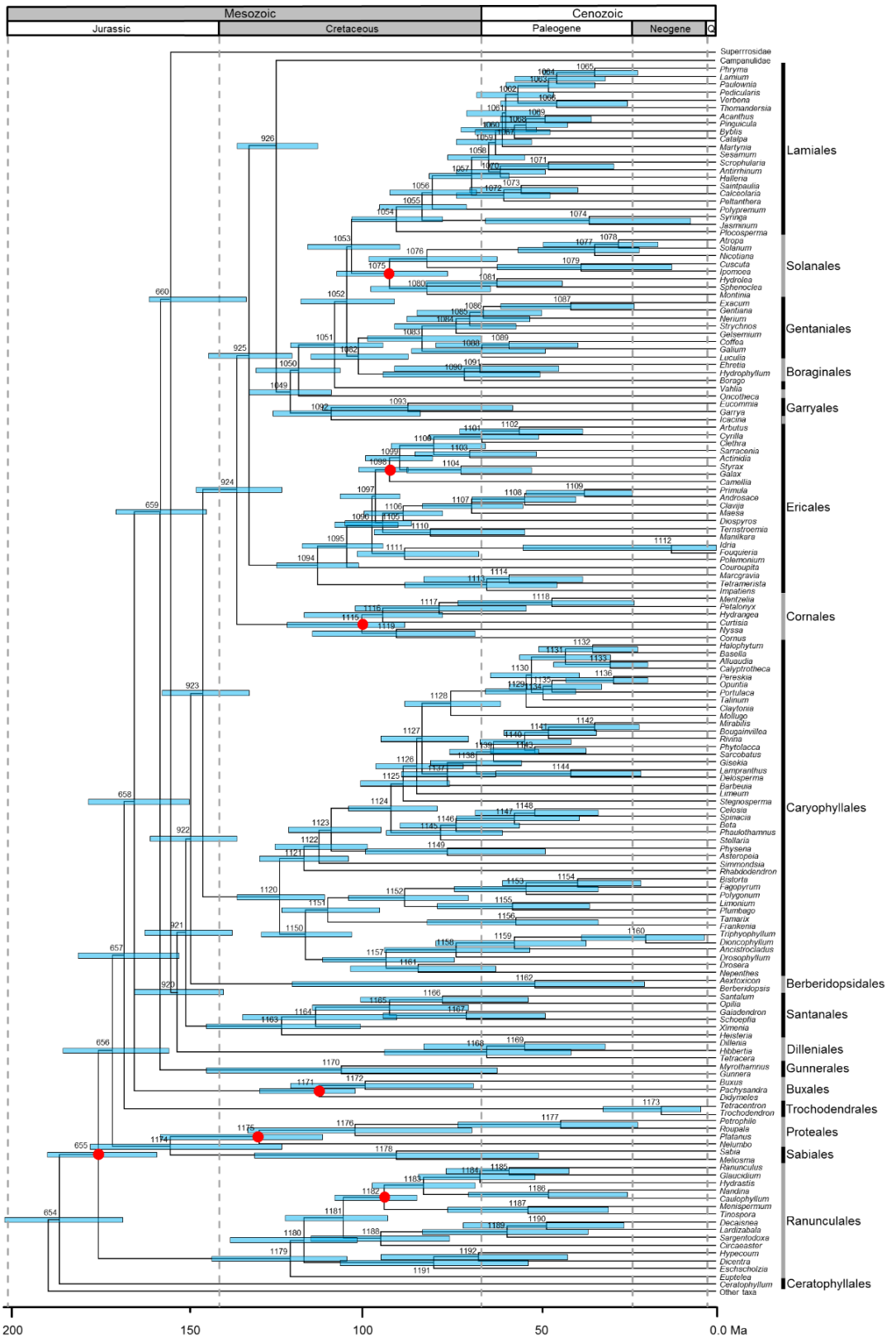


Figure B.2.1: Chronogram of 644 taxa of tracheophytes (from SA-IR-3P). Continued.

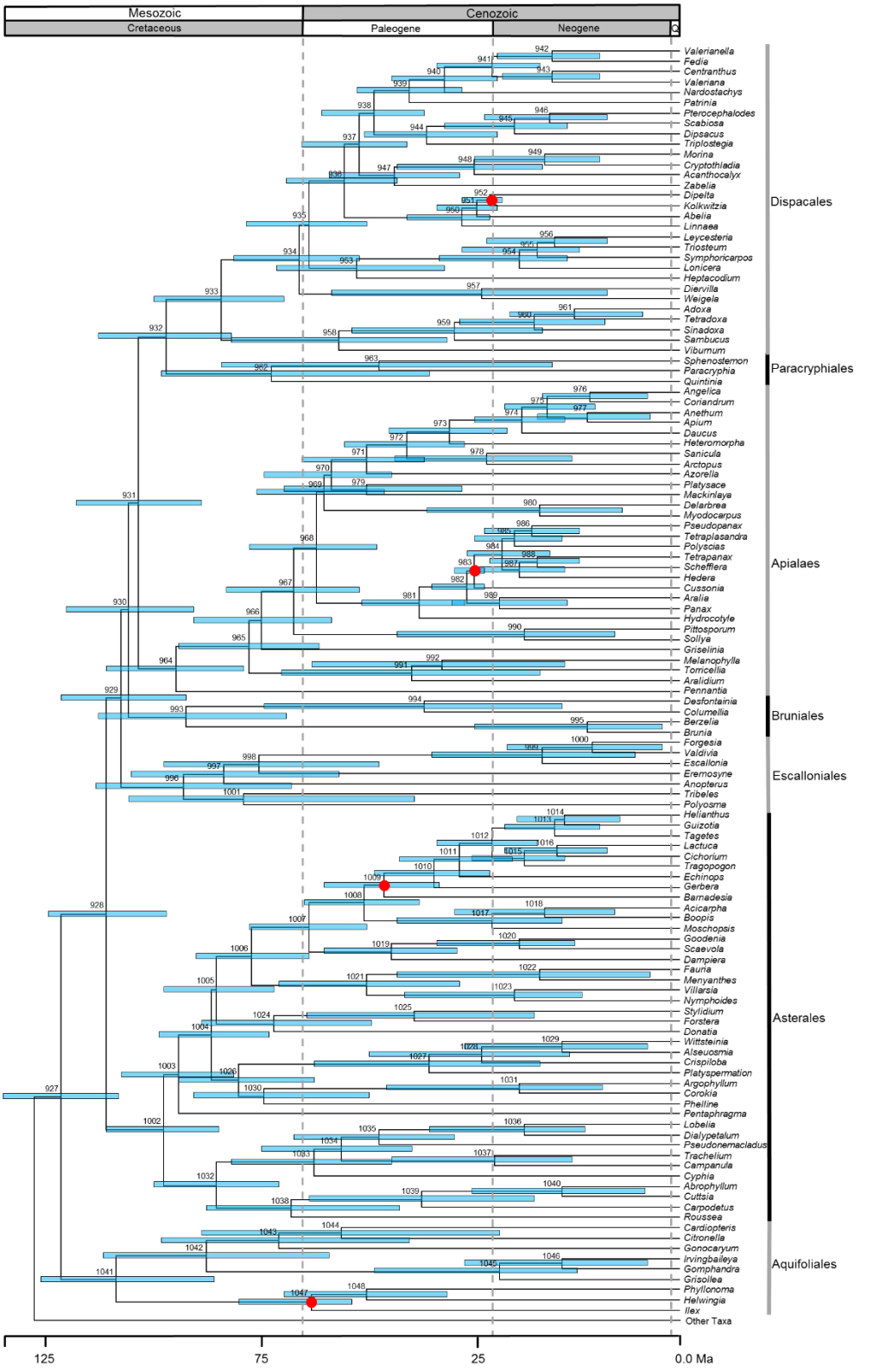


Figure B.2.2: Chronogram of 644 taxa of tracheophytes (from SA-IR-3P). Continued.

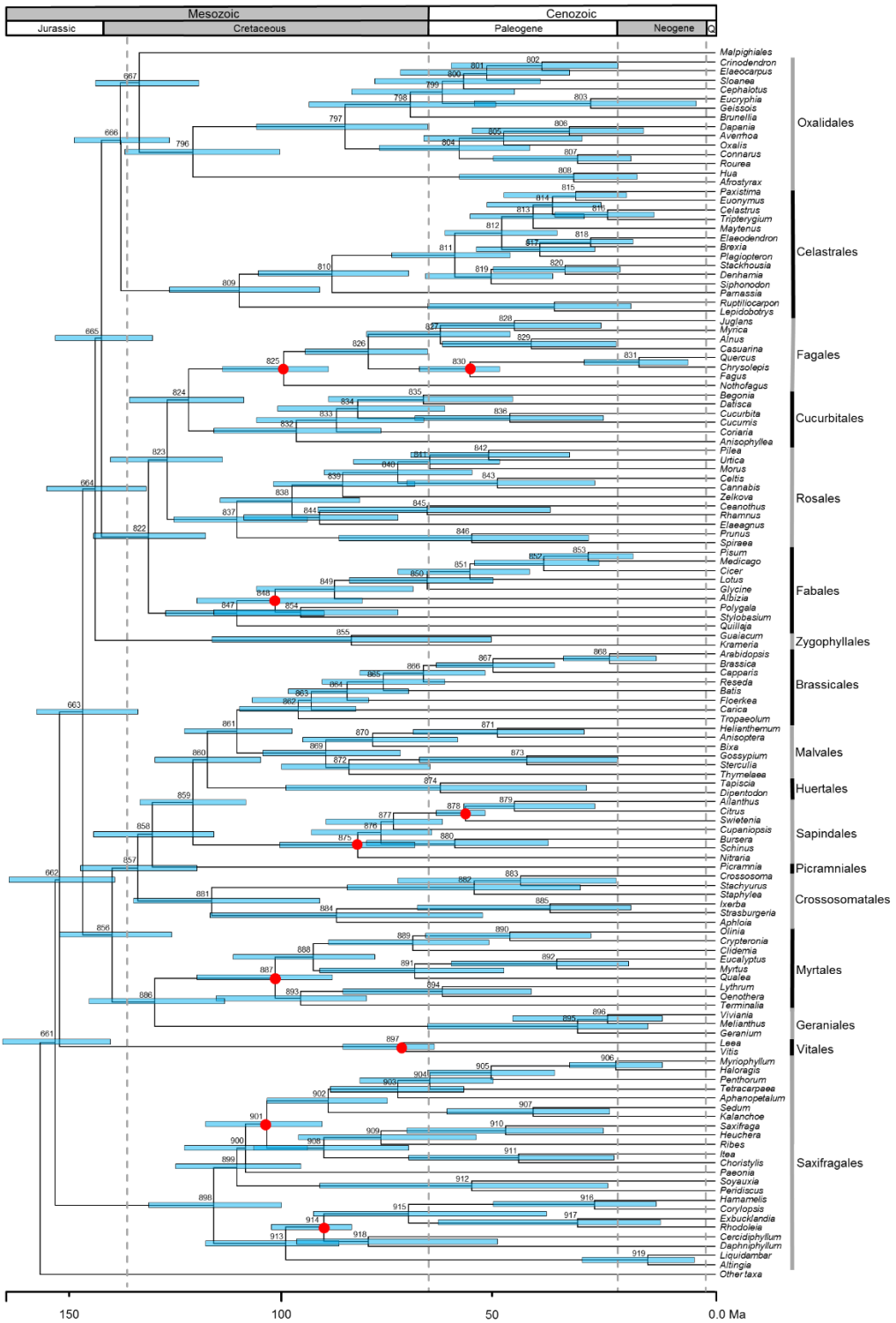


Figure B2.3: Chronogram of 644 taxa of tracheophytes (from SA-IR-3P).

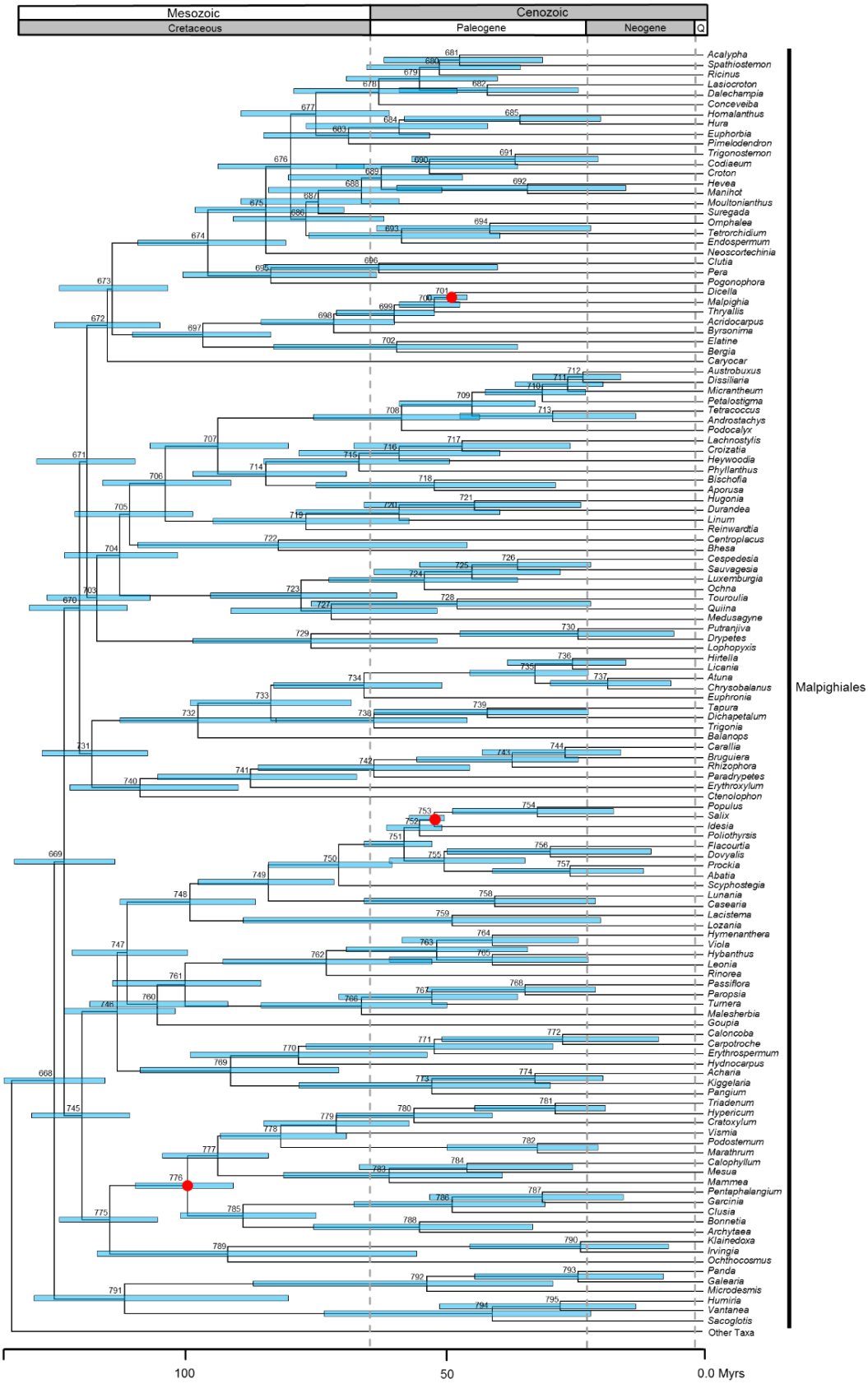


Figure B.2.4: Chronogram of 644 taxa of tracheophytes (from SA-IR-3P). Continued.

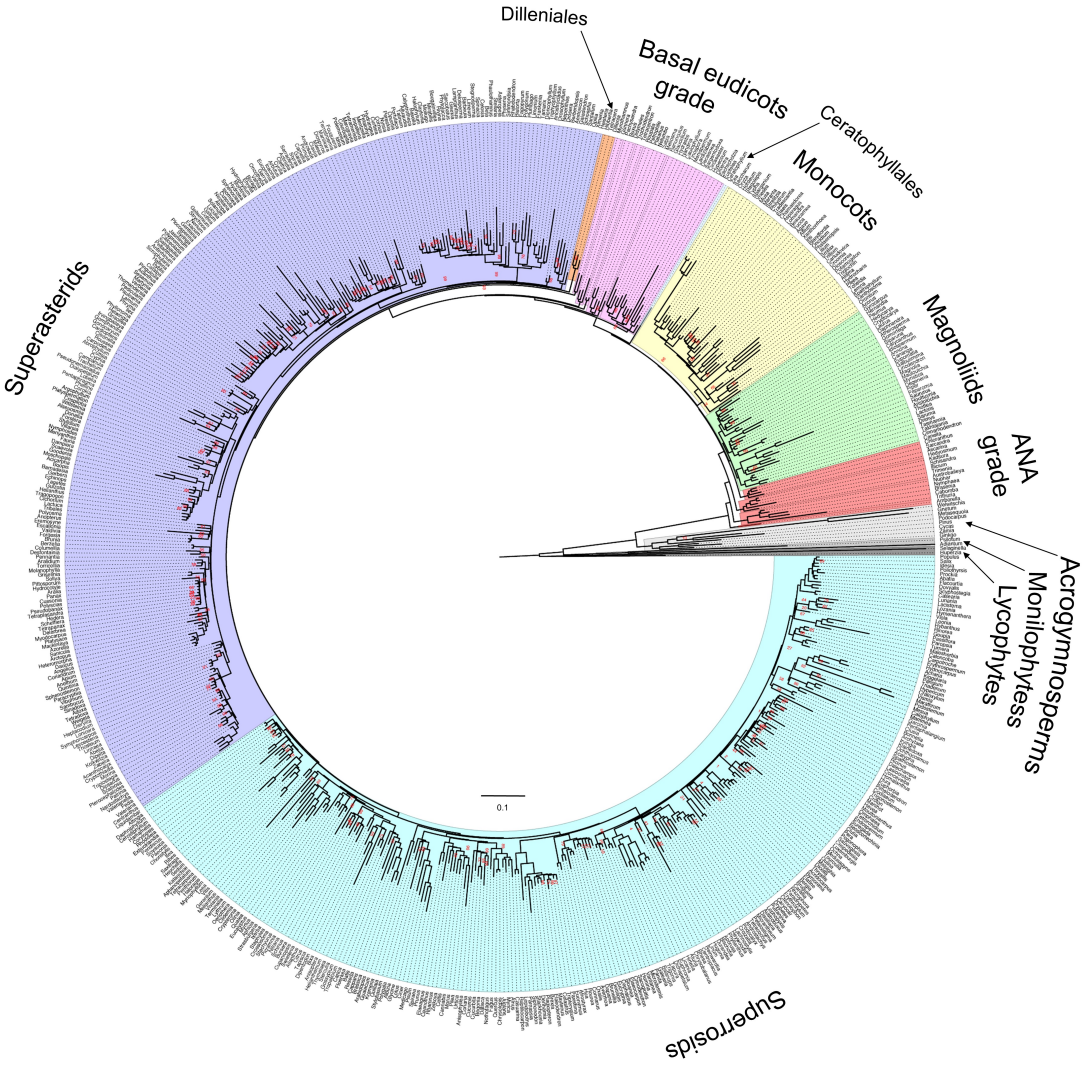


Figure B.3: RAxML phylogenetic tree from 83 genes and 644 taxa of tracheophytes. The Phylogeny was estimated using the GTR+GAMMA model, using 5 partitions (plastid 1st-2nd, plastid 3rd, mitochondrial 1st-2nd, mithochondrial 3rd and nuclear RNA genes). Bootstrap support values of 100% were excluded, while other support values <100% are reported directly with red numbers on nodes. The major lineages of tracheophytes and major groups of angiosperms are highlighted: lycophytes (dark grey) and monilophytes (middle grey), acrogymnosperms (pale grey), ANA grade (red), magnoliids (green), monocots (yellow), Ceratophyllales (pale blue), basal eudicots grade (pink), Dilleniales (orange), superasterids (purple) and superrosids (blue).

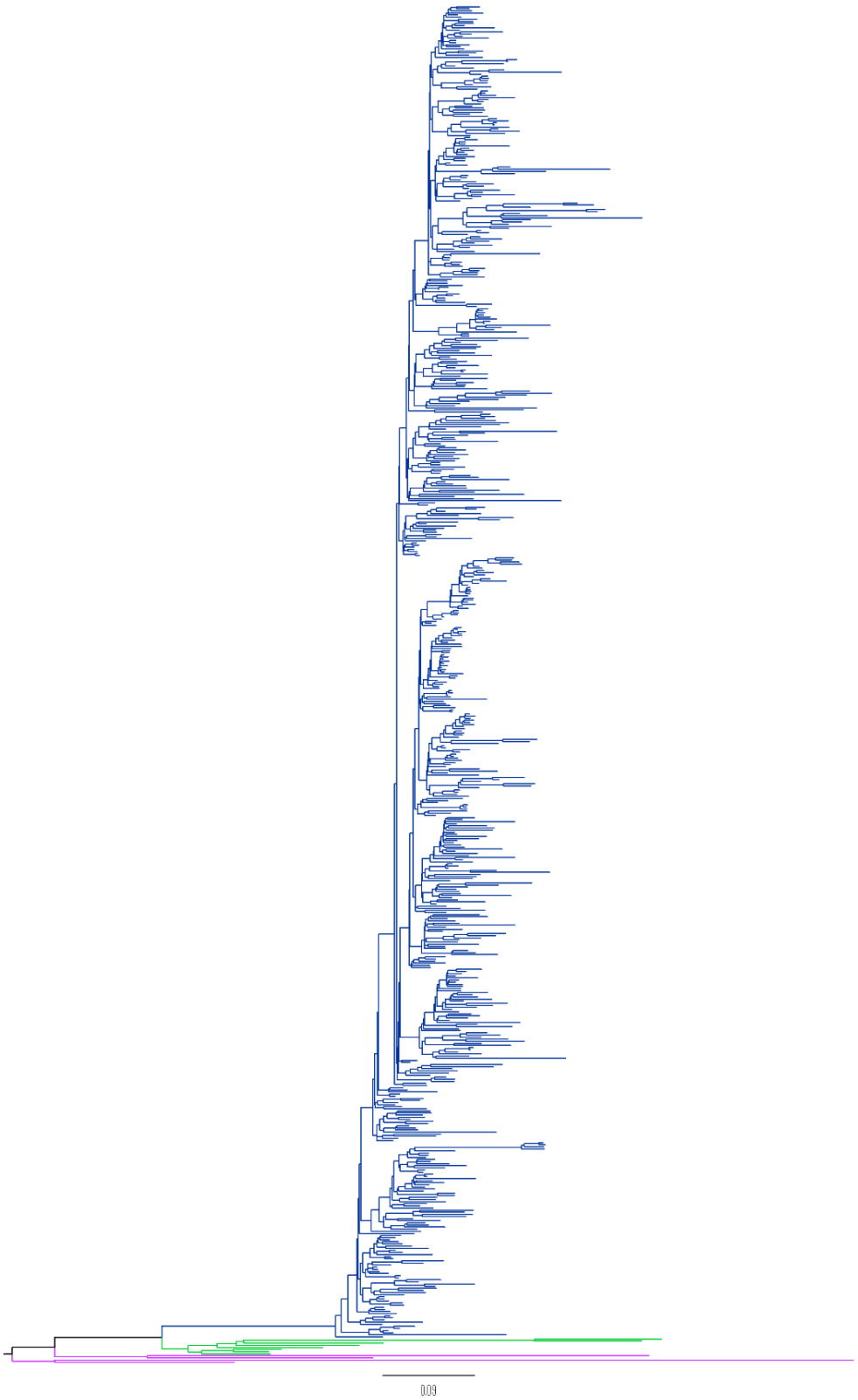


Figure B.4: RAxML phylogenetic tree from plastid 1st-2nd codon positions for 643 taxa
The Phylogeny was estimated using the GTR+GAMMA model and 1 partition (plastid 1st-2nd). The major lineages of tracheophytes are highlighted: lycophytes and monilophytes (purple), acrogymnosperms (green) and angiosperms (blue).

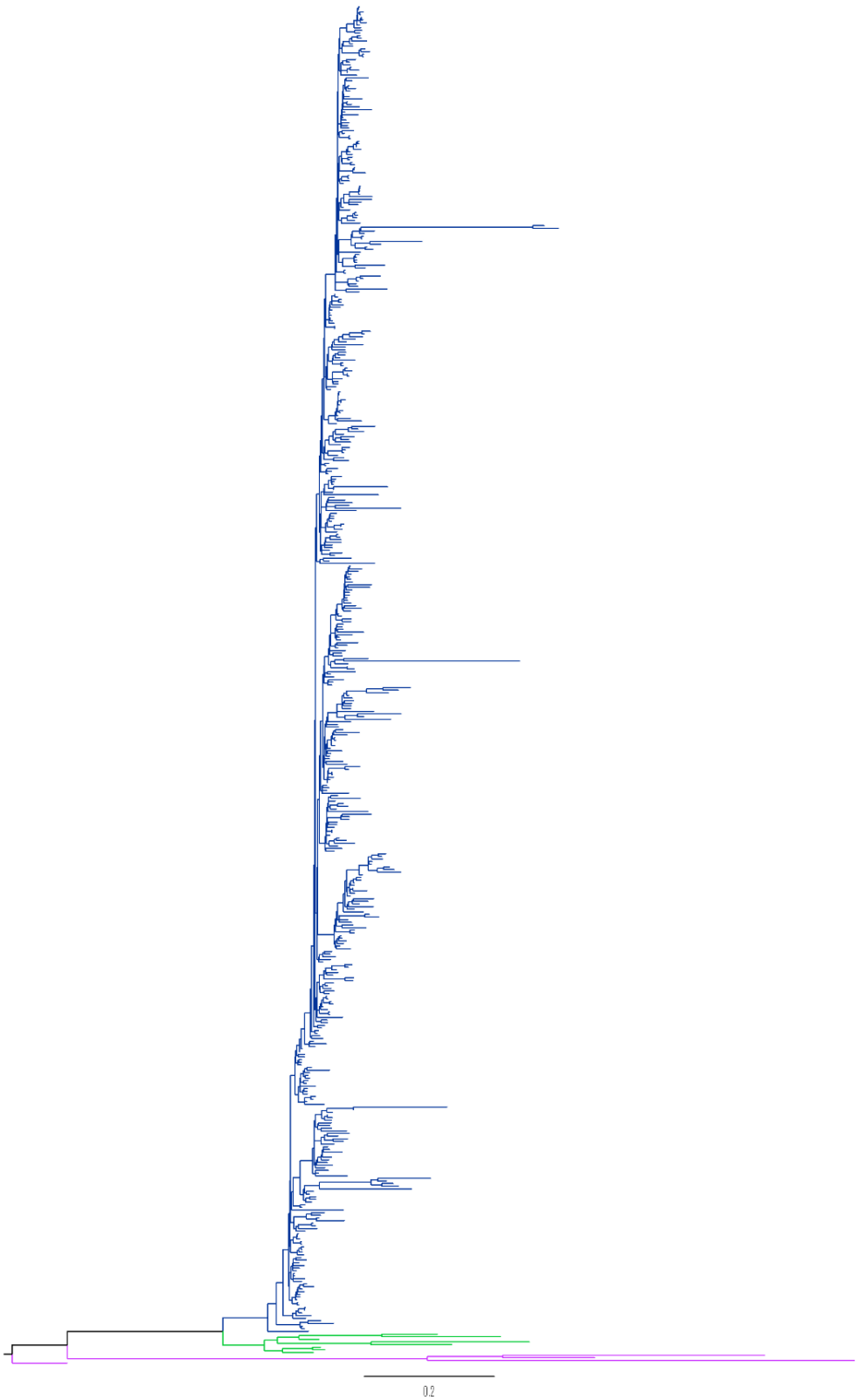


Figure B.5: RAxML phylogenetic tree from mitochondrial 1st-2nd codon positions for 515 taxa. The Phylogeny was estimated using the GTR+GAMMA model and 1 partition (mitochodrial 1st-2ndThe major lineages of tracheophytes are highlighted: lycophytes and monilophytes (purple), acrogymnosperms (green) and angiosperms (blue).

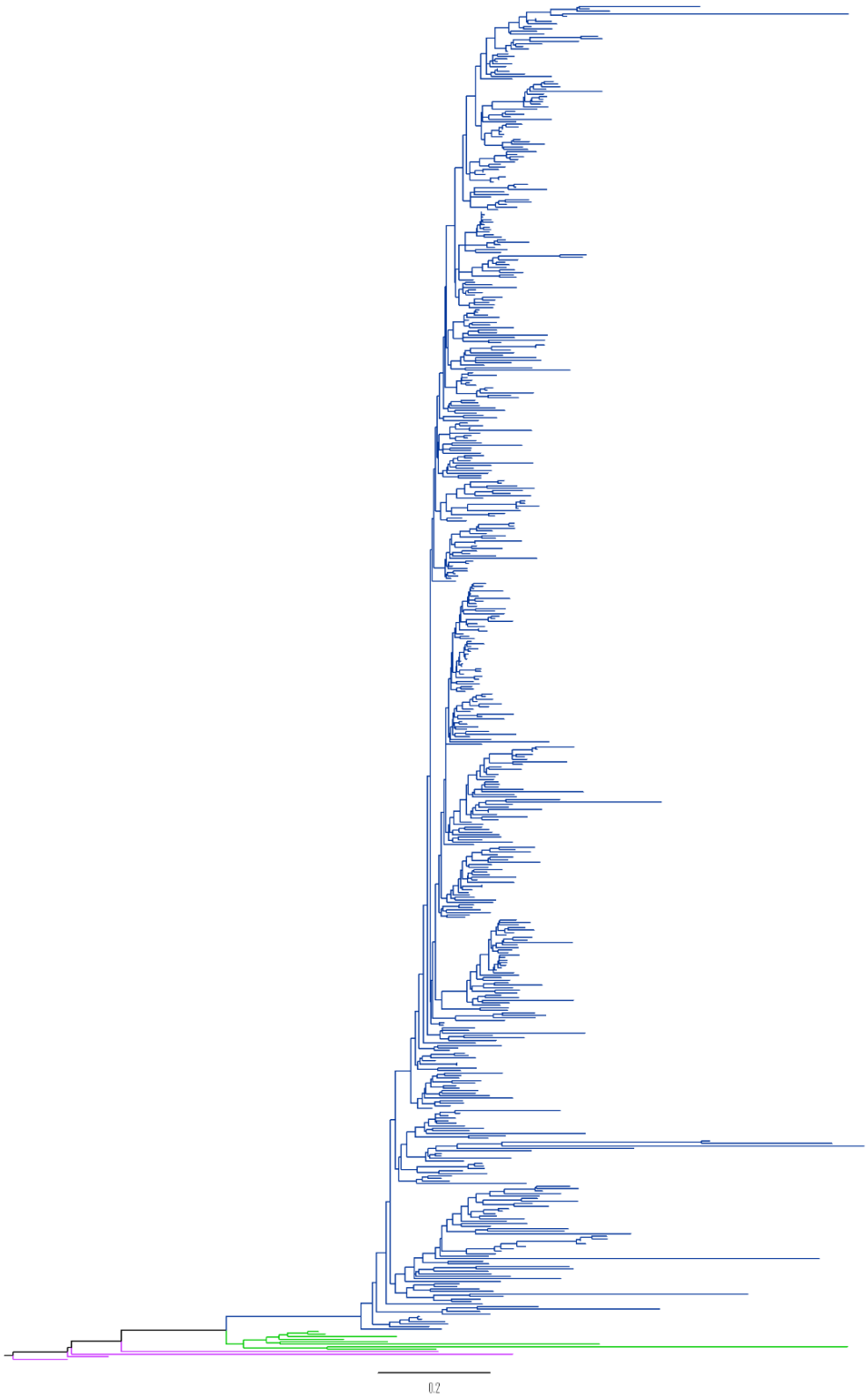
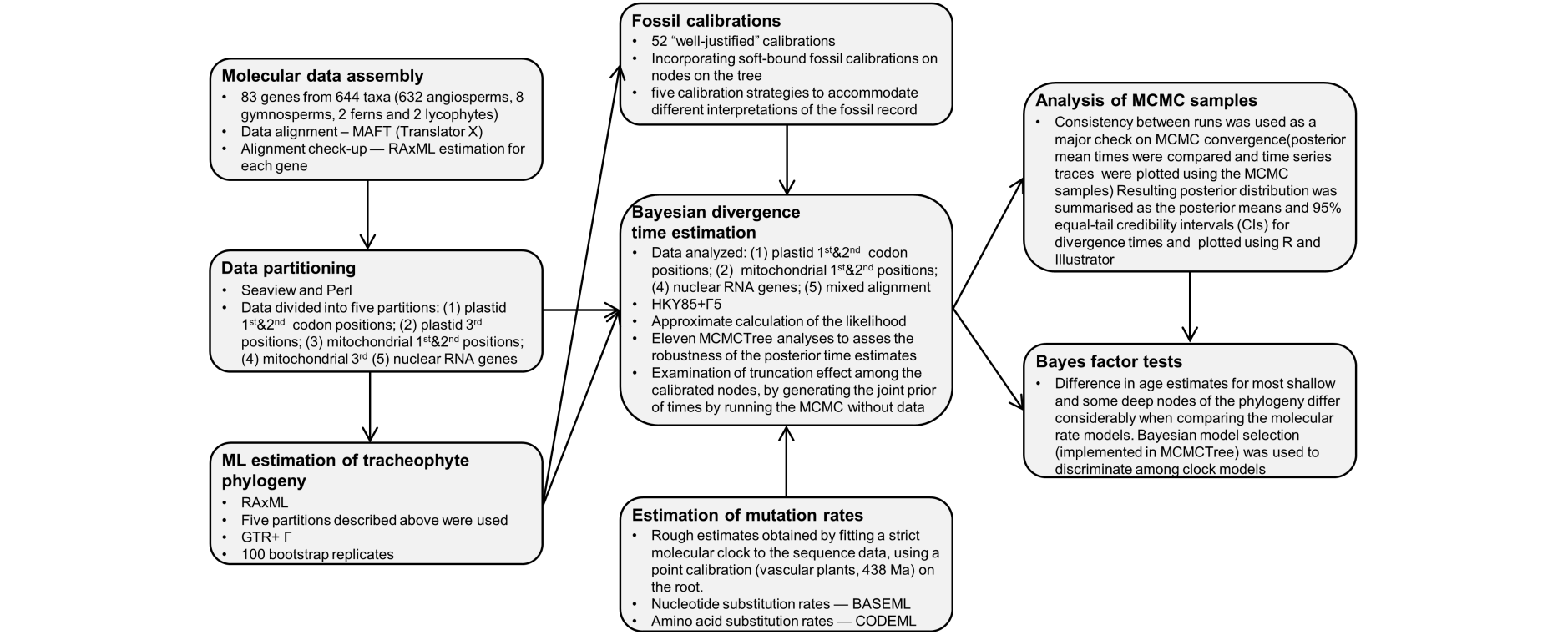


Figure B.6: RAxML phylogenetic tree from nuclear RNA genes for 540 taxa. The Phylogeny was estimated using the GTR+GAMMA model and 1 partition (nuclear RNA genes). The major lineages of tracheophytes are highlighted: lycophytes and monilophytes (purple), acrogymnosperms (green) and angiosperms (blue).



The scripts and datasets used in this workflow are available at: https://figshare.com/s/163a3425d44d2fb482b2 and https://figshare.com/s/404b70bc39656c2cf57e respectively.

Figure B.7: Workflow for chapter three. Scripts: https://figshare.com/s/163a3425d44d2fb482b2; datasets: https://figshare.com/s/404b70bc39656c2cf57e

Table B.1: List of genes included in the dataset. Tree lengths for each gene were obtained using RAxML under GTR+GAMMA model.

gene	No. taxa	# sites	tree length	gene	No. taxa	# sites	tree length
Plastid Genes				Plastid Genes			
accD	83	1200	7.03	*psbJ*	109	120	6.26
atpA	110	1518	6.98	*psbK*	110	126	5.31
atpB	559	1548	21.24	*psbL*	108	111	2.9
atpE	108	387	7.25	*psbM*	107	102	4.47
atpF	105	543	7.49	*psbZ*	106	183	5.06
atpH	110	240	5.62	*rbcL*	627	1497	43.26
atpI	109	741	7.34	*rpl14*	109	363	7.04
ccsA	109	672	10.56	*rpl16*	110	393	8.86
cemA	108	654	7.42	*rpl2*	100	813	4.49
clpP	102	582	12.32	*rpl20*	108	336	9.49
infA	77	222	6.42	*rpl22*	94	318	10.21
matk	588	2421	50.09	*rpl23*	99	270	3.72
ndhA	102	1086	8.97	*rpl32*	102	129	10.96
ndhB	98	1530	3.25	*rpl33*	108	195	7.24
ndhC	100	360	5.76	*rpl36*	110	108	6.58
ndhD	102	1488	9.71	*rpoA*	108	978	9.61
ndhE	103	300	8.19	*rpoB*	108	3207	8.23
ndhF	509	2928	43.81	*rpoC1*	102	2157	8.08
ndhG	101	525	9.22	*rpoC2*	398	8604	28.96
ndhH	101	1173	6.9	*rps11*	109	414	11.43
ndhI	101	465	6.16	*rps12*	104	369	5.28
ndhJ	100	468	5.55	*rps14*	109	294	7.51
ndhK	96	855	7.32	*rps15*	104	255	11.61
petA	108	891	8.72	*rps16*	95	234	7.35
petB	110	642	6.30	*rps18*	107	294	10.83
petD	108	471	6.41	*rps19*	109	273	8.57
petG	110	108	3.99	*rps2*	107	705	7.54
petL	102	183	3.44	*rps4*	370	984	16.63
petN	105	87	2.87	*rps7*	108	465	3.3
psaA	109	2235	6.24	*rps8*	110	390	8.74
psaB	109	2199	5.62	*ycf2*	97	3813	5.91
psaC	109	225	5.68	*ycf3*	108	504	5.74
psaI	107	108	7.15	*ycf4*	106	549	7.57
psaJ	108	123	6.58	Mitochondrial genes			
psbA	109	1056	4.18	*atp1*	378	1713	11.03
psbB/N/T/H	343	2073	16.16	*matr*	480	4236	10.85
psbC	109	1356	6.31	*nad5*	373	2073	5.60
psbD	109	1056	4.65	*rps3*	398	3171	21.05
psbE	110	243	5.12	Nuclear RNA genes			
psbF	109	117	2.97	*rDNA* 18_S	501	1769	29.03
psbI	109	105	5.99	*rDNA* 26_S	368	3396	37.04

Table B.2: Basic information of data partitions.

Partition	No. taxa	# sites	ML tree lenght	alpha
Nucleotide partitions				
Plastid 1st & 2nd	643	41021	22.541	0.519
Plastid 3rd	643	20360	47.605	1.180
Mitochondrial 1st & 2nd	515	5690	13.666	0.473
Mitochondrial 3rd	515	2878	22.82	0.776
Nuclear rDNA	540	5081	39.546	0.372
All partitions concatenated as a single partition	644	75030	29.41	0.486
Amino acid partitions				
Plastid proteins	643	21579	46.563	0.521
Mitochondrial proteins	515	3731	23.003	0.610

Note: Tree length (sum of branch lengths) and gamma shape parameter (α) were estimated using RAxML under the GTR+GAMMA model for nucleotide partitions, the CPREV+GAMMA for plastid proteins and WAG+GAMMA for mitochondrial proteins.

Table B.3: Summary of fossil calibrations used in this study

Node	Clade	Stem/Crown	Minimum divergence time (Ma)	Maximum divergence time (Ma)
645	Tracheophytes	Crown	422 (†*Zosterophyllum* sp)	451 (oldest occurrence of trilete spores)
646	Euphyllophytes	Crown	385.57 (†*Rellimia thomsonii)*	451 (oldest occurrence of trilete spores)
647	Spermatophytes	Crown	308.14 (†*Cordaites iowensis*)	365.63 (base of Vco zone which contains the first seeds)
648	Angiosperms	Crown	125.9 (tricolpate pollen)	247.3 (sediments below the oldest occurrence of angiosperm like pollen which are devoid of such pollen)
651	Mesangiosperms	Crown	125.9 (tricolpate pollen)	247.3 (sediments below the oldest occurrence of angiosperm like pollen which are devoid of such pollen)
655	Eudicots	Crown	119.6 (†*Hyrcantha decussata*)	—
701	*Malphigia*	Stem	44.83 (†*Perisyncolporites pokornyi*)	—
753	*Salix* plus *Populus*	Stem	48.57 (†*Pseudosalix handleyi*)	—
776	Clusiaceae	Stem	85.8 (†*Paleoclusia chevaliery*)	—
825	Fagales	Crown	85.8 (†*Nothofagidites senectus*)	—
830	Fagaceae	Crown	47.6 (†*Fagus langevinii*)	—
848	Polygalaceae	Stem	61.6 (†*Paleosecuridaca curisii*)	—
875	Sapindales	Crown	59.24 (†*Dipteronia brownii*)	—
878	*Alianthus* plus *Citrus*	Stem	51.83 (†*Ailanthus confucii*)	—
887	Myrtales	Crown	83.3 (†*Esqueiria futabensis*)	—
897	Vitales	Crown	65.5 (†*Indovitis chitaleyae)*	—
901	Saxifragales core	Crown	85.8 (†*Divisestylus brevistamineus* and †*D. longistamineus*)	—
914	Hamamelidaceae	Stem	82 (†*Androdecidua endressii*)	—
952	*Dipelta*	Stem	33.7 (†*Diplodipelta reniptera*)	—
983	Araliaceae core	Crown	37.3 (†*Dendropanax eocenensis*)	—
1009	Asteraceae minus *Barnadesia*	Stem	41.5 (†*Tubulifloridites antipodica*)	—
1047	Aquifoliaceae	Stem	61.6 (†*Ilex hercynica*)	—
1075	Solanales	Crown	37.3 (†*Solanites crassus*)	—
1098	Ericales core	Crown	85.8 (†*Paleoenkianthus sayrevillensis*)	—
1115	Cornales	Crown	85.8 (†*Tylerianthus crossmanensis*)	—
1171	Buxales	Crown	100.1 (†*Spanomera marylandensis*)	—
1175	Proteales	Crown	107.59 (†*Sapindopsis variabilis*, †*Aquia brookensis* and †*Palatonocarpus brookensis*)	—
1182	Menispermaceae	Stem	83.41 (†*Prototinomiscium testudinarum* and †*P. vangerowii*)	—
1193	Monocots	Crown	112.6 (†*Liliacidites*)	—
1209	Musaceae	Stem	74.6 (†*Spirematospermum chandlerae*)	—
1211	Arecales	Crown	83.41 (†*Sabalites carolinensis*)	—
1221	Orchidaceae	Crown	17.82 (†*Meliorchis caribea*)	—
1222	Liliales	Crown	18.7 (†*Luzuriaga contortus*)	—
1224	Discorales-Pandanales	Crown	85.8 (†*Mabelia connatifila*)	—
1226	Alismatales	Crown	96.24 (†*Mayoa portugallica*)	—
1228	*Alisma-Potamogeton*	Crown	66 (†*Cardstonia tolmanii*)	—
1232	Araceae	Crown	76 (†*Lysichiton austriacus*)	—
1237	Laurales	Crown	107.59 (†*Virginianthus calycanthoides* and †*Cohongarootonia hispida*)	—
1248	Magnoliales	Crown	110.87 (†*Schenkeriphyllum glanduliferum* and †*Endressinia brasiliana*,)	—
1256	Canellales	Stem	125.9 (†*Walkeripollis gabonensis*)	—
1260	*Saururus*	Stem	44.3 (†*Saururus tuckerae*)	—
1268	Chloroanthaeles	Crown	92.8 (†*Pennipolis sp.*)	—
1272	Schisandraceae	Stem	107.59 (†*Anacostia virginiensis*)	—
1273	*Illicium*	Stem	85.44 (†*Illiciospermum pusillum*)	—
1276	Cabombaceae	Stem	110.97 (†*Pluricarpellatia peltata*)	—
1278	Nymphaeaceae	Crown	92.8 (†*Monetianthus mirus*)	—
1279	Acrogymnospermae	Crown	308.14 (†*Cordaties iowensis*)	365.63 (base of Vco zone which contains the first seeds)
1280	Conifers	Crown	147 (†*Rissikia media*)	312.38 (sediments bearing †*Cordaites iowensis*)
1282	Gnetales	Crown	119.6 (†*Eoantha zherkihinii*)	312.38 (sediments bearing †*Cordaites iowensis*)

2284	*Ginkgo-Cycas*	Crown	264.7 (†Crossozamia)	365.63 (base of Vco zone which contains the first seeds)
1286	Monilophytes	Crown	384.71 (†*Ibyka amphikoma*)	451 (oldest occurrence of trilete spores)
1287	Lycophytes	Crown	392.1 (†*Leclercquia complexa*)	451 (oldest occurrence of trilete spores)

Time unit is 1 Myr. Nodes are numbered as in Figure B.2.

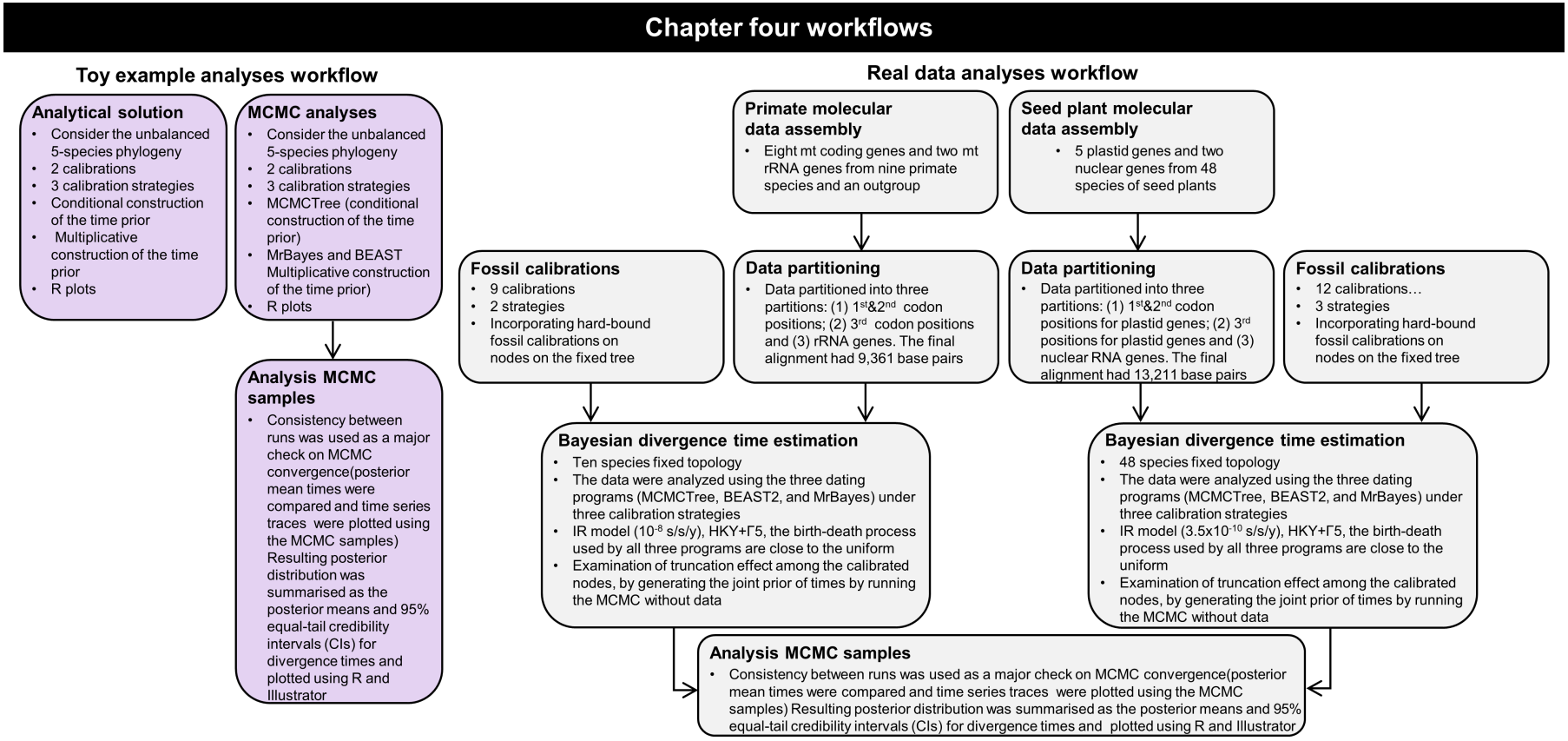
Table B.4:The 95% HPD limits of posterior divergence times, in millions of years before the present, for selected nodes in the vascular plant tree under different prior assumptions (i.e. partition strategies, AR rate model, birth-death parameters and excluding lycophytes and monilophytes)

Node	Clade	SA-IR-1P		SA-IR-MP		SA-AR-3P		SA-IR-3P-EP		SA-IR-3P-BD1		SA-IR-3P-BD2	
645*	Tracheophytes	437	457	443	458	437	456	N/A	N/A	444	460	443	458
646*	Euphylophytes	418	452	430	454	422	451	N/A	N/A	425	454	433	454
647*	Spermatophytes	341	375	355	389	344	371	355	391	354	393	354	418
648*	Angiosperms	223	263	213	261	230	265	231	271	214	263	217	257
651*	Mesangiosperms	184	222	177	211	189	216	186	223	178	214	180	211
655*	Eudicots	164	196	159	187	166	188	166	198	160	190	161	188
661	Superrosids	143	167	140	160	138	154	145	170	139	163	140	160
662	Rosids	141	165	138	158	137	153	143	169	138	161	139	158
668	**Malpighiales**	118	138	114	131	109	121	119	139	115	134	116	132
701*	Stem-*Malphighia*	45	53	44	52	44	50	44	52	44	51	44	52
753*	Stem-*Salix* plus *Populus*	49	58	48	55	49	55	48	55	48	55	48	55
776*	Stem-Clusiaceae	91	115	89	107	85	93	91	112	89	108	89	107
796	**Oxidales**	104	140	105	135	116	130	106	142	102	135	101	134
809	**Celastrales**	95	131	90	123	100	116	95	130	91	124	91	124
825*	**Fagales**	86	111	86	109	94	115	86	112	86	109	86	108
830*	Fagaceae	48	67	48	64	51	80	48	65	48	64	48	65
832	**Cucurbitales**	67	113	78	112	85	110	81	118	78	113	77	112
837	**Rosales**	91	125	91	120	100	118	96	128	92	123	93	121
847	**Fabales**	87	129	89	124	91	116	93	131	89	125	89	124
848*	Stem-*Polygalaceae*	79	123	80	117	86	112	84	124	80	117	81	117
855	**Zygophyllales**	48	124	50	111	66	111	53	117	49	112	50	112
862	**Brassicales**	76	110	82	107	78	100	85	112	82	108	82	107
869	**Malvales**	70	108	74	103	77	98	76	107	73	103	73	103
874	**Huertales**	29	103	31	98	54	101	29	100	30	97	29	97
875*	**Sapindales**	65	99	68	98	77	100	67	100	66	97	67	97
878*	Stem-*Alianthus* plus *Citrus*	52	63	52	62	52	64	52	62	52	62	52	62
881	**Crossosomatales**	83	134	84	131	111	130	95	140	93	134	91	133
887*	**Myrtales**	84	117	86	116	84	104	87	123	85	116	85	113
895	**Geraniales**	85	134	12	58	110	130	10	80	22	41	14	61
897*	**Vitales**	62	84	63	81	63	100	62	81	62	81	62	81
898	**Saxifragales**	101	137	96	126	108	134	99	132	97	127	97	127
901*	Saxifragales core	88	119	87	114	94	122	89	119	88	114	87	114
914*	Stem-Hamamelidaceae	81	101	81	94	89	118	81	98	81	97	81	97
921	Superasterids	141	166	138	159	138	155	143	169	138	161	139	158
924	Asterids	127	151	123	144	126	141	128	153	124	146	124	144
933	**Dispacales**	84	111	76	100	95	111	79	105	75	100	76	100
952*	Stem-Dipelta	34	42	34	40	34	41	33	40	34	40	34	41
962	**Paracryphiales**	38	106	45	97	84	110	51	104	49	100	50	100
964	**Apiales**	89	117	85	110	101	115	87	114	84	110	84	108
983*	Araliaceae core	37	44	37	43	38	59	37	42	37	42	37	42
993	**Bruniales**	77	121	74	111	99	117	79	116	77	112	77	110
996	**Escalloniales**	74	120	79	112	103	118	81	117	78	113	77	111
1002	**Asterales**	97	122	89	109	92	108	91	114	89	109	89	109
1009*	Stem-Asteraceae minus *Barnadesia*	47	77	46	68	53	72	47	71	46	68	46	67
1041	**Aquifoliales**	91	129	90	121	109	126	94	129	92	123	91	121
1047*	Stem-*Aquifoliaceae*	61	83	62	81	66	105	62	83	62	81	62	81
1054	**Lamiales**	82	109	76	100	86	102	81	107	78	103	77	100
1075*	**Solanales**	78	112	77	104	93	109	80	111	79	108	78	105
1083	**Gentaniales**	69	103	65	96	79	99	69	102	67	99	66	96
1090	**Boraginales**	37	95	48	90	77	98	51	96	50	94	49	92

1093	**Garryales**	45	110	55	107	91	118	62	116	59	112	59	111
1094	**Ericales**	103	130	100	123	101	118	102	128	100	123	99	122
1098*	Ericales core	85	104	85	99	86	101	85	100	85	98	85	98
1115*	**Cornales**	86	125	85	114	107	133	85	119	85	118	85	117
1120	**Caryophyllales**	116	143	112	134	106	122	114	139	110	134	111	134
1162	**Berberidopsidales**	18	105	18	94	64	138	18	112	18	111	18	109
1163	**Santanales**	93	145	103	145	117	142	107	151	103	144	103	143
1168	**Dilleniales**	36	109	36	85	37	88	42	95	38	92	39	92
1170	**Gunnerales**	66	162	72	150	104	147	66	151	63	146	61	145
1171*	**Buxales**	99	131	99	124	102	144	99	127	99	125	99	124
1173	**Trochodendrales**	5	55	4	30	18	73	4	31	3	29	3	28
1175*	**Proteales**	108	156	108	148	150	176	108	159	108	152	108	151
1178	**Sabiales**	37	133	47	125	91	154	48	134	48	129	46	129
1179	**Ranunculales**	105	157	102	141	107	156	103	142	101	138	101	139
1182*	Stem-Menispermaceae	83	108	83	106	83	118	83	106	82	103	83	103
1193*	Monocots	141	184	146	179	154	182	150	193	146	182	149	182
1200	**Poales**	76	108	74	105	84	101	79	113	82	108	69	104
1208	**Zingiberales**	75	90	75	87	75	86	75	91	75	90	75	89
1209*	Stem-*Musaceae*	74	86	74	84	74	83	74	85	74	85	74	84
1210	**Commelinales**	47	96	52	96	80	99	57	103	62	99	61	99
1211*	**Arecales**	82	96	82	95	82	90	82	96	82	95	82	95
1212	**Asparagales**	90	124	92	122	105	120	99	130	95	124	94	123
1221*	Orchidaceae	18	23	18	21	18	21	18	22	18	22	18	21
1222*	**Liliales**	51	120	66	114	88	115	69	122	67	117	66	114
1224*	Discorales-Pandanales	87	132	89	126	109	130	92	139	90	132	92	131
1225	**Pandanales**	33	101	45	98	72	109	47	104	44	99	45	99
1226*	**Alismatales**	101	140	105	141	121	147	109	154	106	145	106	144
1228*	*Alisma-Potamogeton*	66	91	66	95	66	83	67	105	67	96	67	96
1232*	Araceae	76	108	76	114	91	130	76	122	76	114	76	115
1235	Magnoliidae	135	188	140	183	179	207	147	201	144	192	144	195
1237*	**Laurales**	108	136	108	133	123	162	108	138	108	135	108	134
1248*	**Magnoliales**	110	134	110	128	117	157	110	133	110	130	110	129
1256*	Stem-Canellales	118	169	127	170	169	201	134	190	130	180	130	186
1257	**Piperales**	96	142	108	148	113	169	111	160	108	154	108	156
1260*	Stem-*Saururus*	44	61	44	62	44	72	44	65	44	63	44	64
1264	**Canellales**	70	137	68	131	124	179	70	140	68	136	68	136
1268*	**Chloranthales**	92	121	92	114	93	144	92	114	92	114	92	113
1271	**Austrobaileyales**	115	175	119	168	132	192	120	174	119	170	119	169
1272*	Stem-Schisandraceae	107	135	107	133	109	153	107	135	107	134	107	133
1273*	Stem-*Illicium*	85	107	85	103	85	114	85	103	85	104	85	103
1275	**Nymphaeales**	124	197	125	185	145	208	132	202	130	193	129	196
1276*	Stem-*Cabombaceae*	110	135	110	128	110	137	110	130	110	129	110	129
1278*	Nymphaceae	93	116	93	116	94	124	93	118	93	116	93	116
1279*	Acrogymnosperms	307	346	307	345	306	326	310	362	307	343	308	355
1280*	Conifers	217	314	265	323	262	309	278	328	268	320	267	328
1282*	Gnetales	116	169	117	164	116	170	116	161	117	156	115	153
1284*	*Ginkgo-Cycas*	263	313	263	305	273	315	264	319	263	312	263	311
1286*	Monilophytes	383	411	383	410	383	405	N/A	N/A	383	398	384	413
1287*	Lycophytes	391	423	390	445	392	441	N/A	N/A	392	447	390	428

Note: Nodes are numbered as in Figure B.2. The 52 calibrated nodes are represented by (*) and nodes in bold characters represent major angiosperm orders. Posterior times are the 95% HPD CI, estimated using the HKY85+Γ_5 substitution model, and calibration strategy A (SA). 1P, the tree partitions analysed as a single partition; MP, mixed partitions of plastid proteins, mitochondrial proteins and nuclear RNA genes; IR, independent rates model; AR autocorrelated rates model and birth-death parameters adjusted to generate a tree with long internal branches and short tip branches (BD1) and large node ages with nodes close to the root (BD2).

C. Supplementary figures and tables for chapter four



The scripts and datasets used in this workflow are available at Figshare: https://figshare.com/s/f57db45a890975837b0a and https://figshare.com/s/2d1ac059646932e74525 respectively.

Figure C.1: Workflow for chapter four. Scripts: https://figshare.com/s/f57db45a890975837b0a; datasets: https://figshare.com/s/2d1ac059646932e74525

Table C.1: Primate fossil calibrations used in this study

Node	Clade	Minimum (Ma)	Maximum (Ma)
11	Scandentia-Primates	61.5 (†*Carpolestidae*)	130 (absence of placentals)
12	Primates (*Otolemur*-Human)	55.6 (†*Altiatlasius*)	—
13	Haplorhini (*Tarsius*-Human)	45 (†*Tarsius*)	—
14	Anthropoidea (*Callithrix*-Human)	33.7 (†*Catopithecus*)	—
15	Catarrhini (*Macaca*-Human)	23.5 (†*Proconsul*)	34 (absence of hominoids)
16	Hominidae (*Pongo*-Human)	11.2 (†*Sivapithecus*)	33.7 (absence of pongines)
17	Ponginae (*Gorilla-Pan*/Human)	7.25 (†*Chororapithecus*)	—
18	Homininae (*Pan*-Human)	5.7 (†*Orrorin*)	10 (absence of hominines)
19	Lorisoidea (*Otolemur-Microcebeus*)	33.7(†*Karanisia*)	55.6 (absence of strepsirrhines)

Note: Calibrations are derived from dos Reis et al. (2012). Nodes are numbered as in Figure 4.4.

Table C.2: Seed plant fossil calibrations used in this study

Node	Clade	Minimum (Ma)	Maximum (Ma)
49	Spermatophyta (*Ginkgo-Quercus*)	308.14 (†*Cordaites iowensis*)	365.63 (base of Vco zone which contains the first seeds)
50	Angiospermae (*Amborella-Quercus*)	125.9 (tricolpate pollen)	247.3 (sediments below the oldest occurrence of angiosperm like pollen which are devoid of such pollen)
57	Eudicotyledonae without *Ceratophyllum* (*Nandina-Quercus)*	119.6 (†*Hyrcantha decussata*)	—
65*	No name *(Arabidopsis-Quercus)*	82.8 (†*Paleoclusia chevalieri* & †*Dressiantha bicarpellata*)	127.2 (oldest potential age of tricolpate pollen)
70	Vitales (*Vitis-Leea*)	65.6 (†*Indovitis chitaleyae)*	—
76	Cornales (*Petalonix-Cornus*)	85.8 (†*Tylerianthus crossmanensis*)	—
77	Proteales (*Nelumbo-Platanus*)	107.59 (†*Sapindopsis variabilis*, †*Aquia brookensis* and †*Palatonocarpus brookensis*)	—
78	Total group Monocotyledones (*Acorus-Musa*)	112.6 (†*Liliacidites*)	—
84	Chloroanthales (*Chloranthus-Hedyosmum*)	92.8 (†*Pennipolis*)	—
86	No name (*Trimenia-Illicium*)	107.59 (†*Anacostia virginiensis*)	—
88	Cabombaceae (*Cabomba-Nymphaea*)	111 (†*Pluricarpellatia peltata*)	—
89	Acrogymnospermae (*Ginkgo-Pinus*)	308.14 (†*Cordaties iowensis*)	365.7 (base of Vco zone which contains the first seeds)
90	Coniferae (*Pinus-Metasequoia*)	147 (†*Rissikia media*)	312.38 (sediments bearing †*Cordaites iowensis*)
92	Gnetales (*Gnetum-Welwitschia*)	119.6	312.38 (sediments bearing †*Cordaites iowensis*)
94	No name (*Ginkgo-Cycas*)	264.7 (†Crossozamia)	365.63 (base of Vco zone which contains the first seeds)

Note: Calibrations are derived from Barba-Montoya et al. (Unpublished results) and (*) from Clarke et al. (2011). Nodes are numbered as in Figure 4.4.

Table C.3: GenBank accession numbers of genes included in the primate dataset

Sampled taxa	Mitochondrial protein coding								Mitochondrial rRNA	
	Cyt B	*CO1*	*CO2*	*CO3*	*ND2*	*ND3*	*ND4*	*ND4L*	*12S*	*16S*
Callithrix jacchus	AB572419	AB572419	AB572419	AB572419	AB572419	AB572419	AB572419	AB572419	AB572419	AB572419
Gorilla gorilla	NC_011120	NC_011120	NC_011120	NC_011120	NC_011120	NC_011120	NC_011120	NC_011120	NC_011120	NC_011120
Homo sapiens	NC_012920	NC_012920	NC_012920	NC_012920	NC_012920	NC_012920	NC_012920	NC_012920	JQ724861	NC_012920
Macaca mulatta	NC_005943	NC_005943	NC_005943	NC_005943	NC_005943	NC_005943	NC_005943	NC_005943	NC_005943	NC_005943
Microcebus murinus	GU327180	EU179510	GU326994	AF224624	—	AF224624	AF224624	AF224624	AY582694	AF072424
Otolemur garnettii	AY441466	AY671787	—	—	—	—	—	—	DQ073511	AF072430
Pan troglodytes	NC_001643	NC_001643	NC_001643	NC_001643	NC_001643	NC_001643	NC_001643	NC_001643	NC_001643	NC_001643
Pongo abelii	NC_002083	NC_002083	NC_002083	NC_002083	NC_002083	NC_002083	NC_002083	NC_002083	NC_002083	NC_002083
Tarsius syrichta	AB371090	AB371090	AB371090	AB371090	AB371090	AB371090	AB371090	AB371090	AB371090	AB371090
Tupaia belangeri	NP_065227	NP_065217	NP_065218	NP_065221	NP_065216	NP_065222.1	NP_065224.1	NP_065223	NC_002521	NC_002521

Table C.4: GenBank accession numbers of genes included in the seed plant dataset.

Sampled taxa	Plastid protein coding					Nuclear rRNA	
	atpB	*matK*	*NdhF*	*rbcL*	*rps4*	*18s*	*26s*
Acorus	156622714	69217282	62903246	37959585	69216101	1280175	2687430
Amborella	—	77743619	12005302	37194760	32401792	1777635	30527312
Arabidopsis	NC_000932	NC_000932	NC_000932	NC_000932	—	255689731	—
Austrobaileya	—	77743621	9623112	37194768	AF313613	1022919	30527315
Berberidopsis	HQ843255	EU002171	EU002201	EU002274	EU002295	AF206866	AF389242
Buxus	5001578	33333425	13491703	81230666	—	1369761	22595016
Cabomba	—	33333427	11022839	336459	—	470806	30527316
Calycanthus	AJ235422	AF543730	AF123802	L14291	AY832286	U38318	30527317
Ceratophyllum	—	33333433	6424767	6513623	—	2588926	30527319
Chloranthus	—	—	85740637	37194780	16565398	470845	19919686
Cornus	AY725918	37935834	170178386	85678969	157689453	18108	85678959
Cycas	156597988	15866114	156597953	156598062	16565388	470889	66969255
Daucus	113200887	2281160	113200887	1374996	113200887	—	37778850
Dioscorea	17224736	10863023	—	17224611	6002044	194022474	18032178
Ginkgo	69214415	170320099	156598259	459408	125662705	471060	30527338
Gnetum	—	42529088	—	34733649	16565390	471074	2687428
Gunnera	EU002162	AM396506	157689283	11323502	157689481	1777743	22595023
Hedyosmum	—	89242559	85740639	HQ336536	—	7595445	30527324
Ilex	GQ997300	EF590403	22796548	EF590536	GQ983972	7595458	19919644
Illicium	—	33333445	11022853	37194806	125662731	471848	66969247
Ipomoea	EU118126	EU118126	EU118126	EU118126	EU118126	1049331	6707928
Leea	AJ235520	AF274621	157689307	AJ235783	157689509	AY674612	9799451
Liquidambar	157689159	AF015651	157689309	86373225	157689513	471888	19919658
Liriodendron	—	7239757	6424765	13539638	—	471902	30527327
Magnolia	—	18025023	16416698	18024760	32401803	471957	22595029
Metasequoia	8439471	9279988	33416153	4049493	125662697	125662660	66969251
Morus	8439474	30421073	33950066	533039	—	532608	19919673
Musa	156598338	GQ374866	156598303	342515	156598367	125661879	27462207

Nandina	904133	23495303	24934989	904135	32401801	904137	22595014
Nelumbo	194267386	AM396514	193957791	229464449	GQ997619	472018	22595032
Nymphaea	—	77743669	39598867	342748	32401794	472389	30527328
Oryza	AB037543	7140883	—	344016	—	472190	—
Oxalis	257853489	EU002186	157689327	257783291	157689531	AF206978	19919671
Petalonyx	6689009	21464757	6706314	10945630	JF268455	6688995	30230598
Pinus	—	148832433	—	86559793	16565394	472085	20467279
Platanus	194267388	166156329	193957793	86373193	32401796	472132	9799460
Podocarpus	33318668	259191248	33416155	33317783	32401837	20502	66969252
Populus	AF209658	EU749357	AY757172	EU676964	—	DQ371807	19919559
Quercus	157689165	206604156	157689339	46091778	157689545	37729422	37993790
Spinacia	AF528861	NC_002202	AY090621	AJ400848	AJ400848	L24420	HQ843464
Staphylea	157689167	157689209	84872975	57490179	157689559	—	19919574
Trimenia	—	89242601	37544941	7580491	—	—	30527333
Trithuria	21684894	—	33333926	118152381	—	—	—
Trochodendron	157689169	2149797	157689369	7240475	157689575	—	19919646
Vitis	DQ424856	AF274635	DQ424856	L01960	DQ424856	AF207053	AF479207
Welwitschia	—	14579066	—	4049566	32401829	472502	110757094
Yucca	156622714	AB088789	33333914	37722389	69216111	61741965	—
Zamia	—	15866124	33416161	32811545	32401756	475140	66969254

References

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* **38**: W7-13.

Adams K. 2013. Genomic clues on the ancestral flowering plant. *Science* **342**: 1456-1457.

Adamson RS. 1939. The clasification of life-forms of plants. *Botanical Review* **5**: 546-561.

Amborella_Genome_Project. 2013. The Amborella genome and the evolution of flowering plants. *Science* **342**: 1241089.

APG_IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**: 1-20.

Augusto L, Davies TJ, Delzon S, De Schrijver A. 2014. The enigma of the rise of angiosperms: can we untie the knot? *Ecology Letters* **17**: 1326-1338.

Barba-Montoya J, Dos Reis M, Schneider H, Donoghue PCJ, Yang Z. Submitted 2017. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytologist*.

Battistuzzi FU, Filipski AJ, Kumar S. 2011. Molecular clock: testing. In. *eLS*: John Wiley & Sons, Ltd: Chischester.

Beaulieu JM, O'meara BC, Crane P, Donoghue MJ. 2015. Heterogeneus rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. *Systematic Biology* **64**: 869-878.

Bell CD. 2015. Between a Rock and a Hard Place: Applications of the "Molecular Clock" in Systematic Biology. *Systematic Botany* **40**: 6-13.

Bell CD, Soltis DE, Soltis PS. 2005. The age of the angiosperms: a molecular timescale without a clock. *Evolution* **59**: 1245-1258.

Bell CD, Soltis DE, Soltis PS. 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* **97**: 1296-1303.

Benson RBJ, Mannion PD, Butler RJ, Upchurch P, Goswami A, Evans SE. 2013. Cretaceous tetrapod fossil record sampling and faunal turnover: implications for biogeography and the rise of modern clades. *Palaeogeography, Palaeoclimatology, Palaeoecology* **372**: 88-107.

Benton MJ. 2010. The origins of modern biodiversity on land. *Philosophical Transactions of the Royal Society B* **365**: 3667-3679.

Benton MJ, Donoghue PCJ, Asher RJ. 2009. Calibrating and constraining molecular clocks. In: Hedges SB and Kumar S, editors. *The timetree of life*. Oxford, UK: Oxford University Press. p. 35-86.

Bouckaert R, J., Heled D, Kuhnert T, Vaughan CH, Wu DX, Suchard MA, Rambaut A, J. DA. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*

Bowe LM, Coat G, dePamphilis CW. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 4092–4097.

Bracken-Grissom HD, Ahyong ST, Wilkinson RD, Feldmann RM, Schweitzer CE, Breinholt JW, Bendall M, Palero F, Chan T-Y, Felder DL et al. 2014. The emergence of lobsters: phylogenetic relationships, morphological evolution and divergence time comparisons of an ancient group (Decapoda: Achelata, Astacidea, Glypheidea, Polychelida). *Systematic Biology* **63**: 457-479.

Brenner GJ. 1996. Evidence for the earliest stage of angiosperm pollen evolution: a paleoequatorial section from Israel. In: Taylor DW and Hickey LJ, editors. *Flowering Plant Origin, Evolution & Phylogeny*. New York: Chapman & Hall. p. 91–115.

Brenner GJ, Bickoff IS. 1992. Palynology and age of the Cretaceous Basal Kurnub Group from the Coastal Plain to the Northern Negev of Israel. *Palynology* **16**: 137-185.

Bromham L, Penny D. 2003. The modern molecular clock. *Nature Review Genetics* **4**: 216-224.

Brown JW, Smith SA. 2017. The past sure is tense: on interpreting phylogenetic divergence time estimates. *Systematic Biology*.

Cantino PD, Soltis DE, Doyle JA, Soltis PS, Graham SW, Donoghue MJ. 2007. Towards a phylogenetic nomenclature of Tracheophyta. *Taxon* **56**: 822-846.

Cardinal S, Danforth BN. 2013. Bees diversified in the age of eudicots. *Proceedings of the Royal Society B* **280**: 20122686.

Cascales-Minana B, Cleal CJ, P. G. 2016. Is Darwin's 'Abominable Mystery' still a mystery today? *Cretaceous Research* **61**: 256-262.

Chang SC, Zhang HC, Renne PR, Fang Y. 2009. High-precision Ar-40/Ar-39 age for the Jehol Biota. *Palaeogeography, Palaeoclimatology, Palaeoecology* **280**: 94-104.

Christenhusz MJM, Byng JW. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* **261**: 201-217.

Clarke JT, Warnock RC, Donoghue PC. 2011. Establishing a time-scale for plant evolution. *New Phytologist* **192**: 266-301.

Coen ES, Meyerowitz EM. 1991. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**.

Coiffard C, Gomez B, Daviero-Gomez V, Dilcher DL. 2012. Rise to dominance of angiosperm pioneers in European Cretaceous environments. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 20955-20959.

Cornet B. 1989. The reproductive morphology and biology of Sanmiguelia lewissi, and its bearing on angiosperm evolution in the Late Triassic. *Evolutionary Trends in Plants* **3**: 25-51.

Crane PR. 1985. Phylogenetic analysis of seed plants and the origin of angiosperms. *Annals of the Missouri Botanical Garden* **72**: 716-793.

Crane PR. 1987. Review of Cornet, B. The Leaf Venation and Reproductive Structures of a Late Triassic Angiosperm. Sanmiguelia lewisii, Evolutionary Theory 7: 231-309 (1986). *Taxon* **36**: 778-779.

Crane PR, Friis EM, Pedersen KJ. 1995. The origin and early diversification of angiosperms. *Nature* **374**: 27-33.

Crepet WL. 2008. The Fossil Record of Angiosperms: Requiem or Renaissance?1. *Annals of the Missouri Botanical Garden* **95**: 3-33.

Crepet WL, Niklas KJ. 2009. Darwin's second 'abominable mystery': Why are there so many angiosperm species? *American Journal of Botany* **96**: 366-381.

Curtius K, Wong CJ, Hazelton WD, Kaz AM, Chak A, Willis JE, Grady WM, Luebeck EG. 2016. A Molecular Clock Infers Heterogeneous Tissue Age Among Patients with Barrett's Esophagus. *PLoS Computational Biology* **12**: e1004919.

Davies TJ, Barraclough TG, Chase MW, Soltis PS, Soltis DE, Savolainen V. 2004. Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 1904-1909.

DeGroot MH, Schervish MJ. 2002. *Probability and Statistics*. Boston, USA: Addison-Wesley.

Dickerson RE. 1971. The structure of cytochrome c and the rates of evolution. *Journal of Molecular Evolution* **1**: 26-45.

Dilcher D. 2000. Toward a new synthesis: Major evolutionary trends in the angiosperm fossil record. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 7030-7036.

Donoghue PC, Yang Z. 2016. The evolution of methods for establishing evolutionary timescales. *Philosophical Transactions of the Royal Society B* **371**.

Doolittle RF, Blomback B. 1964. Amino-acid sequence investigations of fibrinopeptides from various mammals - evolutionary implications. *Nature* **202**: 147-152.

Dornburg A, Beaulieu JM, Oliver JC, Near TJ. 2011. Integrating fossil preservation biases in the selection of calibrations for molecular divergence time estimation. *Systematic Biology* **60**: 519-527.

dos Reis M, Donoghue PC, Yang Z. 2016. Bayesian molecular clock dating of species divergences in the genomics era. *Nature Review Genetics* **17**: 71-80.

dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PC, Yang Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings of the Royal Society B* **279**: 3491-3500.

dos Reis M, Thawornwattana Y, Angelis K, Telford MJ, Donoghue PC, Yang Z. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Current Biology* **25**: 2939-2950.

dos Reis M, Yang Z. 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Molecular Biology and Evolution* **28**: 2161-2172.

dos Reis M, Yang Z. 2013. The unbearable uncertainty of Bayesian divergence time estimation. *Journal of Systematic Evolution* **51**: 30-43.

dos Reis M, Zhu, T., and Yang, Z. 2014. The impact of the rate prior on Bayesian estimation of divergence times with multiple Loci. *Systematic Biology* **63**: 555-565.

Doyle JA. 2005. Early evolution of angiosperm pollen as inferred from molecular and morphological phylogenetic analyses. *Grana* **44**: 227-251.

Doyle JA. 2006. Seed ferns and the origin of angiosperms. *The Journal of the Torrey Botanical Society* **133**: 169-209.

Doyle JA. 2007. Systematic value and evolution of leaf architecture across the angiosperms in light of molecular phylogenetic analyses. *CFS Courier Forschungsinstitut Senckenberg* **258**: 21-37.

Doyle JA. 2008. Integrating Molecular Phylogenetic and Paleobotanical Evidence on Origin of the Flower. *International Journal of Plant Sciences* **169**: 816-843.

Doyle JA. 2009. Evolutionary significance of granular exine structure in the light of phylogenetic analyses. *Review of Palaeobotany and Palynology* **156**: 198-210.

Doyle JA. 2012. Molecular and Fossil Evidence on the Origin of Angiosperms. *Annual Review of Earth and Planetary Sciences* **40**: 301-326.

Doyle JA, Donoghue MJ. 1986. Seed plant phylogeny and the origin of the angiosperms: an experimental cladistic approach. *Botanical Review* **52**.

Doyle JA, Donoghue MJ. 1993. Phylogenies and angiosperm diversification. *Paleobiology* **19**: 147-167.

Doyle JA, Donoghue MJ, Zimmer EA. 1994. Integration of morphological and ribosomal RNA data on the origin of angiosperms. *Annals of the Missouri Botanical Garden* **81**: 419-450.

Doyle JA, Hotton CL. 1991. Diversification of early angiosperm pollen in a cladistic context. In: Blackmore S and Barnes SH, editors. *Pollen and spores: patterns of divrsification* Oxford: Clarendon Press.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: e88.

Drummond AJ, Stadler T. 2016. Bayesian phylogenetic estimation of fossil ages. *Philosophical Transactions of the Royal Society B* **371**.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evoution* **29**: 1969-1973.

Du Rietz GE. 1931. Life-forms of terrestrial flowering plants. *Acta Phytogeographica Suecica 3* **1**: 95.

Duchene S, Lanfear R, Ho SY. 2014. The impact of calibration and clock-model choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution* **78**: 277-289.

Endress PK, Doyle JA. 2009. Reconstructing the ancestral angiosperm flower and its initial specializations. *American Journal of Botany* **96**: 22-66.

Feild TS, Chatelet DS, Brodribb TJ. 2009. Ancestral xerophobia: a hypothesis on the whole plant ecophysiology of early angiosperms. *Geobiology* **7**.

Feldberg K, Schneider H, Stadler T, Schafer-Verwimp A, Schmidt AR, Heinrichs J. 2014. Epiphytic leafy liverworts diversified in angiosperm-dominated forests. *Scientific Reports* **4**: 5974.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**: 368-376.

Fitch W. 1976. Molecular evolutionary clocks. In: Ayala FJ, editor. *Molecular Evolution*. Sinauer Associates: Sunderland, MA.

Forest F. 2009. Calibrating the Tree of Life: fossils, molecules and evolutionary timescales. *Annals of Botany* **104**: 789-794.

Foster SP, Sauquet H, van der Merve M, McPherson H, Rossetto M, Ho SY. 2016. Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Systematic Biology*.

Friedman WE. 2009. The meaning of Darwin's 'Abominable Mystery'. *American Journal of Botany* **96**: 5-21.

Friis EM, Crane PR, Pedersen KR. 2011. *Early Flowers and Angiosperm Evolution*. Cambridge UK: Cambridge University Press.

Friis EM, Doyle JA, Endress PK, Leng Q. 2003. Archaefructus – angiosperm precursor or specialized early angiosperm? *Trends in Plant Science* **8**: 369-373.

Friis EM, Pedersen KR, Crane PR. 2000a. Fossil floral structures of a basal angiosperm with monocolpate, reticulate-acolumellate pollen from the Early Cretaceous of Portugal. *Grana* **39**: 226-239.

Friis EM, Pedersen KR, Crane PR. 2000b. Reproductive structure and organization of basal angiosperms from the early cretaceous (barremian or aptian) of western Portugal. *International Journal of Plant Sciences* **161**: 5119-5182.

Friis EM, Pedersen KR, Crane PR. 2010. Diversity in obscurity: fossil flowers and the early history of angiosperms. *Philosophical Transactions of the Royal Society B* **365**: 369-382.

Frohlich MW, Chase MW. 2007. After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* **450**: 1184-1189.

Gandolfo MA, Nixon KC, Crepet WL. 2008. Selection of Fossils for Calibration of Molecular Dating Models1. *Annals of the Missouri Botanical Garden* **95**: 34-42.

Gernhard T. 2008. The conditioned reconstructed process. *Journal of Theoretical Biology* **253**: 769-778.

Gomez B, Daviero-Gomez V, Coiffard C, Martin-Closas C, Dilcher DL. 2015. Montsechia, an ancient aquatic angiosperm. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 10985-10988.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27**: 221-224.

Gradstein FM, Ogg JG, Schmitz M. 2012. *The Geologic Time Scale 2012*. Waltham, MA: Elsevier.

Graur D, Li WH. 2000. *Fundamentals of Molecular Evolution*. Sunderland, Massachusetts: Sinauer Associates, Inc.

Guindon S. 2013. From trajectories to averages: an improved description of the heterogeneity of substitution rates along lineages. *Systematic Biology* **62**: 22-34.

Hasegawa M, Kishino H, Yano T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA *Journal of Molecular Evolution* **22**: 160-174.

Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97-108.

Heath TA. 2012. A hierarchical Bayesian model for calibrating estimates of species divergence times. *Systematic Biology* **61**: 793-809.

Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States of America* **111**: E2957-2966.

Heled J, Drummond AJ. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology* **61**: 138-149.

Heled J, Drummond AJ. 2015. Calibrated birth-death phylogenetic time-tree priors for bayesian inference. *Systematic Biology* **64**: 369-383.

Herendeen PS, Friis EM, Pedersen KR, Crane PR. 2017. Palaeobotanical redux: revisiting the age of the angiosperms. *Nature Plants* **3**: 17015.

Hickey LJ. 1997. Early Cretaceous fossil evidence for angiosperm evolution. *Botanical Review* **43**: 1-102.

Hill CR, Crane PR. 1982. Evolutionary cladistics and the origin of angiosperms. In: Joysey KA and Friday AE, editors. *Problems of Phylogenetic Reconstruction*. London: Academic Press.

Hilton J, Bateman RM. 2006. Pteridosperms are the backbone of seed-plant phylogeny. *The Journal of the Torrey Botanical Society* **133**: 119-168.

Himmelmann L, Metzler D. 2009. TreeTime: an extensible C++ software package for Bayesian phylogeny reconstruction with time-calibration. *Bioinformatics* **25**: 2440-2441.

Ho SY. 2014. The changing face of the molecular evolutionary clock. *Trends in Ecology and Evolution* **29**: 496-503.

Ho SY, Duchene S. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology* **23**: 5947-5965.

Ho SY, Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* **58**: 367-380.

Hochuli PA, Feist-Burkhardt S. 2013. Angiosperm-like pollen and Afropollis from the Middle Triassic (Anisian) of the Germanic Basin (Northern Switzerland). *Frontiers in Plant Science* **4**: 344.

Hohna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F. 2016. RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. *Systematic Biology* **65**: 726-736.

Hohna S, Stadler T, Ronquist F, Britton T. 2011. Inferring speciation and extinction rates under different sampling schemes. *Molecular Biology and Evolution* **28**: 2577-2589.

Huges NF. 1994. *The Enigma of Agiosperm Origins*. Cambridge, UK: Cambridge University Press.

Huges NF, McDougall AG. 1987. Records of angiospermid pollen entry into the English early Cretaceous succession. *Review of Palaeobotany and Palynology* **50**: 255-272.

Inoue J, Donoghue PC, Yang Z. 2010. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Systematic Biology* **59**: 74-89.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97-100.

Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ. 2002. *Plant Systematics: A Phylogenetic Approach*. Sunderland, MA: Sinauer Associates.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772-780.

Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* **217**: 264-266.

Kimura M. 1983. The Neutral Theory of Molecular Evolution. In. *Molecular evolutionary rates contrasted with phenotypic evolutionary rates*: Cambridge University Press. p. 55-97.

King CE, Jukes TH. 1969. Non-Darwinian evolution. *Science* **164**: 768-798.

Kishino H, Thorne JL, Bruno WJ. 2001. Performance of a Divergence Time Estimation Method under a Probabilistic Model of Rate Evolution. *Molecular Biology and Evolution* **18**: 352–361.

Laenen B, Shaw B, Schneider H, Goffinet B, Paradis E, Desamore A, Heinrichs J, Villarreal JC, Gradstein SR, McDaniel SF et al. 2014. Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. *Nature Communications* **5**: 5134.

Langley CH, Fitch WM. 1974. An examination of the constancy of the rate of molecular evolution. *Journal of Molecular Evolution* **3**: 161-177.

Larget B. 2010. Introduction to Markov Chain Monte Carlo methods in molecular evolution. In: Nielsen R, editor. *Sttistical methods in molecular evolution*. USA: Springer.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**: 2286-2288.

Lartillot N, Philippe H. 2006. Computing Bayes factors using thermodynamic integration. *Systematic Biology* **55**: 195-207.

Lartillot N, Phillips MJ, Ronquist F. 2016. A mixed relaxed clock model. *Philosophical Transactions of the Royal Society B* **371**.

Lepage T, Bryant D, Philippe H, Lartillot N. 2007. A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution* **24**: 2669-2680.

Lewis PO. 2011. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**: 913-925.

Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, Barker MS. 2015. Early genome duplications in conifers and other seed plants. *Science Advances* **1**.

Linder M, Britton T, Sennblad B. 2011. Evaluation of Bayesian models of substitution rate evolution--parental guidance versus mutual independence. *Systematic Biology* **60**: 329-342.

Liu Z-J, Wang X. 2016a. Yuhania: a unique angiosperm from the Middle Jurassic of Inner Mongolia, China. *Historical Biology*: 1-11.

Liu ZJ, Wang X. 2016b. A perfect flower from the Jurassic of China. *Historical Biology* **28**: 707-719.

Lloyd GT, Davis KE, Pisani D, Tarver JE, Ruta M, Sakamoto M, Hone DW, Jennings R, Benton MJ. 2008. Dinosaurs and the Cretaceous Terrestrial Revolution. *Proceedings of the Royal Society B* **275**: 2483-2490.

Magallón S. 2004. Dating lineages: molecular and paleontological approaches to the temporal framework of clades. *International Journal of Plant Sciences* **165(4 Suppl.)**: S7-S21.

Magallón S. 2009. Flowering plants. In: Hedges SB and Kumar S, editors. *The Timtree of Life*. Oxford UK: Oxford Umiversity Press. p. 161-165.

Magallón S. 2010. Using fossils to break long branches in molecular dating: a comparison of relaxed clocks applied to the origin of angiosperms. *Systematic Biology* **59**: 384-399.

Magallón S. 2014. A Review of the Effect of Relaxed Clock Method, Long Branches, Genes, and Calibrations in the Estimation of Angiosperm Age. *Botanical Sciences* **92**: 1.

Magallón S, Castillo A. 2009. Angiosperm diversification through time. *American Journal of Botany* **96**: 349-365.

Magallón S, Gomez-Acevedo S, Sanchez-Reyes LL, Hernandez-Hernandez T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**: 437-453.

Magallón S, Hilu KW, Quandt D. 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany* **100**: 556-573.

Magallón S, Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution* **55**: 1762-1780.

Magallón S, Sanderson MJ. 2002. Relationships among seed plants inferred from highly conserved genes: sorting conflicting phylogenetic signals among ancient lineages. *American Journal of Botany* **89**: 1991-2006.

Margoliash E. 1963. Primary structure and evolution of cytochrome c. *Proceedings of the National Academy of Sciences of the United States of America* **50**: 672-679.

Marshall CR. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *The American Naturalist* **171**: 726-742.

Martin W, Lydiate D, Brinkmann H, Forkmann G, Seadler H, Cerff R. 1993. Molecular phylogenies in angiosperm evolution. *Molecular Biology and Evolution* **10**: 140-162.

Meredith RW, Janecka JE, Gatesy j, Oliver A. Ryder, Colleen A. Fisher, Emma C. Teeling, Alisha Goodbla, Eduardo Eizirik, Taiz L. L. Simão, Tanja Stadler et al. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**: 521–524.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH. 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**: 1087-1092.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**: 763-767.

Molak M, Suchard MA, Ho SY, Beilman DW, Shapiro B. 2015. Empirical calibrated radiocarbon sampler: a tool for incorporating radiocarbon-date and calibration error into Bayesian phylogenetic analyses of ancient DNA. *Molecular Ecology Resources* **15**: 81-86.

Murat F, Armero A, Pont C, Klopp C, Salse J. 2017. Reconstructing the genome of the most recent common ancestor of flowering plants. *Nature Genetics* **49**: 490-496.

Muse S, Weir B. 1992. Testing for equality of evolutionary rates. *Genetics* **132**.

Near TJ, Bolnick DI, Wainwright P. 2005. Fossil calibrations and molecular divergence time estimates in centrachid fishes (Telostei: Centrarchidae). *Evolution* **59**: 1768-1782.

Nei N, Kumar S. 2000. *Molecular Evolution and Phylogenetics*. New York: Oxford Unoversity Press.

Nixon KC, Crepet WL, Stevenson D, Friis EM. 1994. A reevaluation of seed plant phylogeny. *Annals of the Missouri Botanical Garden* **81**: 484-533.

Nowak MD, Smith AB, Simpson C, Zwickl DJ. 2013. A simple method for estimating informative node age priors for the fossil calibration of molecular divergence time analyses. *PLoS One* **8**: e66245.

O'Reilly JE, dos Reis M, Donoghue PC. 2015. Dating Tips for Divergence-Time Estimation. *Trends in Genetics* **31**: 637-650.

O'Reilly JE, Puttick MN, Parry L, Tanner AR, Tarver JE, Fleming J, Pisani D, Donoghue PC. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters* **12**: 20160081.

Ohta T, Kimura M. 1971. On the constancy of the evolutionary rate in cistrons. *J Mol Evol* **1**: 18-25.

Paradis E. 2013. Molecular dating of phylogenies by likelihood methods: a comparison of models and a new information criterion. *Molecular Phylogenetics and Evolution* **67**: 436-444.

Parham JF, Donoghue PC, Bell CJ, Calway TD, Head JJ, Holroyd PA, Inoue JG, Irmis RB, Joyce WG, Ksepka DT et al. 2012. Best practices for justifying fossil calibrations. *Systematic Biology* **61**: 346-359.

Paton A, Brummitt N, Govaerts R, Harman K, Hinchcliffe S, Allkin B, Lughadha E. 2008. Towards Target 1 of the Global Strategy for Plant Conservation: a working list of all known plant species—progress and prospects. *Taxon* **57**: 602-611.

Piskur J, Rozpedowska E, Polakova S, Merico A, Compagno C. 2006. How did Saccharomyces evolve to become a good brewer? *Trends in Genetics* **22**: 183-186.

Puttick MN, O'Reilly JE, Tanner AR, Fleming J, Clark J, Holloway L, Lozano-Fernandez J, Parry L, Terver JE, Pisani D et al. 2017. Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society B* **284**.

Pyron RA. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology* **60**: 466-481.

Rai HS, O'Brien HE, Reeves PA, Olmstead RG, Graham SW. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular PHylogenetics and Evolution* **29**: 350-359.

Rambaut A, Bromham L. 1998. Estimating divergence dates from molecular sequences. *Molecular Biology and Evolution* **15**: 442-448.

Ramshaw JA, Richardson DL, Meatyard BT, Brown RH, Richardson M, Thompson EW, Boulter D. 1972. Time of origin of flowering plants by using amino-acid sequence data of cytochrome-c. *New Phytologist* **71**: 773-779.

Rannala B, Yang Z. 2007. Inferring speciation times under an episodic molecular clock. *Systematic Biology* **56**: 453-466.

Rannala B, Yang Z. 2013. Molecular clock dating. In: Losos JB, editor. *The Princeton Guide to Evolution*. NJ, USA: Princeton University Press. p. 67-74.

Rannala B, Yang Z. 2017. Efficient Bayesian species tree inference under the multispecies coalescent. *Systematic Biology*: 1-20.

Raven PH, Axelrod DI. 1974. Angiosperm biography and past continental movements. *Annals of the Missouri Botanical Garden* **61**.

Reyes-Garcia V, Huanca T, Valdez V, Leonard W, Wilkie D. 2005. Cultural, practical, and economic value of wild plants: a quantitative study in the Bolivian Amazon. *Economic Botany* **60**: 62-74.

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. 2012a. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Systematic Biology* **61**: 973-999.

Ronquist F, Lartillot N, Phillips MJ. 2016. Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society B* **371**.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012b. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539-542.

Rothwell GW, Serbet R. 1994. Lignophyte phylogeny and the evolution of spermatophytes: a numerical cladistic analysis. *Systematic Biology* **19**: 443-482.

Roure B, Baurain D, Philippe H. 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution* **30**: 197-214.

Ruhfel BR, Gitzendanner MA, Soltis P, Soltis DE, G. B. 2014. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* **14**: 1-26.

Rutschmann F. 2006. Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. *Diversity* **12**: 35-48.

Rydin C, Källersjö M, Friis E. 2002. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems, and the monophyly of conifers. *International Journal of Plant Sciences* **163**: 197-214.

Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**: 1218-1231.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19**: 101-109.

Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**: 301-302.

Sanderson MJ, Doyle JA. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from rbcl and 18s rdna data. *American Journal of Botany* **88**: 1499–1516.

Sarich VM, Wilson AC. 1967a. Inmunological time scale for hominid evolution. *Science* **158**: 1200-1203.

Sarich VM, Wilson AC. 1967b. Rates of albumin evolution in primates. *Proceedings of the National Academy of Science of the United States of America* **58**: 142-148.

Sarich VM, Wilson AC. 1973. Generation time and genomic evolution in primates. *Science* **179**: 1144-1147.

Sauquet H, Ho SY, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, Bayly MJ, Bromham L, Brown GK, Carpenter RJ et al. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of Nothofagus (Fagales). *Systematic Biology* **61**: 289-313.

Schneider H, Schuettpelz E, Pryer K, Cranfill R, Magallon S, Lupia R. 2004. Ferns diversified in the shadow of angiosperms. *Nature* **428**: 553–557.

Simpson MG. 2006. *Plant Systematics*. London UK: Elsevier Academic Press.

Smith SA, Beaulieu JM, Donoghue MJ. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 5897-5902.

Smith SA, O'Meara BC. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**: 2689-2690.

Soltis DE, Bell CD, Kim S, Soltis PS. 2008. Origin and early evolution of angiosperms. *Annals of the New York Academy of Sciences* **1133**: 3-25.

Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlsward BS et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *Am J Bot* **98**: 704-730.

Soltis DE, Soltis PS, Zanis MJ. 2002. Phylogeny of seed plants based on evidence from eight genes. *American Journal of Botany* **89**: 1670-1681.

Soltis PS, Soltis DE. 2004. The orgin and diversification of angiosperms. *American Journal of Botany* **91**: 1614–1626.

Soltis PS, Soltis DE, Chase MW, Endress PK, Crane PR. 2004. The diversification of flowering plants. In: Cracraft J and J. DM, editors. *The tree of life*. Oxford, UK: Oxford University Press.

Stadler T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* **261**: 58-66.

Stadler T. 2010. Sampling-through-time in birth-death trees. *Journal of Theoretical Biology* **267**: 396-404.

Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**: 456-463.

Stigler SM. 1986. *The History of Statistics : The Measurement of Uncertainty Before 1900*. Cambridge, Mass, London: Belknap Press of Harvard University Press.

Sun G, Ji Q, Dilcher DL, Zheng S, Nixon KC, Wang X. 2002. Archaefructaceae, a New Basal Angiosperm Family. *Science* **296**: 988-904.

Tajima F. 1993. Simple methods for testing molecular clock hypothesis. *Genetics* **135**: 599-607.

Takezaki N, Rzhetsky A, Nei M. 1995. Phylogenetic test of the molecular clock and linearized trees *Molecular Biology and Evolution* **12**: 823-833.

Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. 2012. Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 19333-19338.

Tavaré S, R. MC, Will O, Soligos C, D MR. 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature* **416**: 726-729.

Taylor DW, Hickey LJ. 1996. Evidence for and implications of an herbaceous origin for angiosperms. In: Taylor DW and Hickey LJ, editors. *Flowering Plant Origin, Evolution & Phylogeny*. New York: Chapman & Hall. p. 232–266.

Taylor EL, Taylor TN. 2009. Seed ferns from the late Paleozoic and Mesozoic: Any angiosperm ancestors lurking there? *American Journal of Botany* **96**: 237-251.

Thorne JL, Kishino H. 2010. Estimation of divergence times from molecular sequence data. In: Nielsen R, editor. *Statistical methods in molecular evolution*. USA: Springer. p. 233-256.

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* **15**: 1647-1657.

Warnock RC, Parham JF, Joyce WG, Lyson TR, Donoghue PC. 2015. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proceedings of the Royal Society B* **282**: 20141013.

Warnock RC, Yang Z, Donoghue PC. 2012. Exploring uncertainty in the calibration of the molecular clock. *Biol Lett* **8**: 156-159.

Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *American Journal of Botany* **102**: 1753-1756.

Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America* **111**: E4859-4868.

Wilkinson RD, Steiper ME, Soligo C, Martin RD, Yang Z, Tavare S. 2011. Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic Biology* **60**: 16-31.

Wright AM, Hillis DM. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS ONE* **9,**.

Wu C-I, Li W-H. 1985. Evidence of higher rates nucleotide substitution in rothents than in man. *Proceedings of the National Academy of Sciences of the United States of America* **82**: 1741-1745.

Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* **60**: 150-160.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* **39**: 306-314.

Yang Z. 2004. A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. *Acta Zoologica Sinica* **50**: 645-656.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586-1591.

Yang Z. 2014. *Molecular Evolution: A Statistical Approach*. Oxford: Oxford University Press.

Yang Z. 2016. Bayesian Phylogenetic Methods. In: Kliman RM, editor. *Encyclopedia of Evolutionary Biology*: Oxford: Academic Press. p. 137-140.

Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo Method. *Molecular Biology and Evolution* **14**: 717-724.

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* **23**: 212-226.

Yang Z, Rannala B. 2012. Molecular phylogenetics: principles and practice. *Nature Review Genetics* **13**: 303-314.

Yang Z, Yoder AD. 2003. Comparison of Likelihood and Bayesian Methods for Estimating Divergence Times Using Multiple Gene Loci and Calibration Points, with Application to a Radiation of Cute-Looking Mouse Lemur Species. *Systematic Biology* **52**: 705-716.

Yoder AD, Yang Z. 2000. Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution* **17**: 1081-1090.

Zahn LM, Kong H, Leebens-Mack JH, Kim S, Soltis PS, Landherr LL, Soltis DE, Depamphilis CW, Ma H. 2005. The evolution of the SEPALLATA subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* **169**: 2209-2223.

Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB et al. 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**: 89-92.

Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications* **5**: 4956.

Zhang C, Stadler T, Klopfstein S, Heath TA, Ronquist F. 2016. Total-evidence dating under the fossilized birth-death procesS. *Systematic Biology* **65**.

Zheng Y, Wiens JJ. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution* **94**: 537-547.

Zhu T, dos Reis M, Yang Z. 2015. Characterization of the uncertainty of divergence time estimation under relaxed molecular clock models using multiple loci. *Systematic Biology* **64**: 267-280.

Zuckerkandl E, Pauling L. 1962. Molecular Disease, Evolution, and Genetic Heterogeneity. In: Kasha M and Pullman B, editors. *Horizons in Biochemistry*. New York: Academic Press. p. 189-225.

Zuckerkandl E, Pauling L. 1965. Evolving genes and proteins. In: Bryson V and Vogel HJ, editors. *Evolutionary Divergence and Convergence in Proteins*. New York: Academic Press. p. 97-166.